Doctoral theses at NTNU, 2022:10

Michele Conni

Appearance Characterization of Textiles

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Michele Conni

Appearance Characterization of Textiles

Thesis for the Degree of Philosophiae Doctor

Gjøvik, January 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

ISBN 978-82-326-5381-2 (printed ver.)
ISBN 978-82-326-6733-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2022:10

Printed by NTNU Grafisk senter

Abstract

The appearance of an object is the combination of the physical attributes that influence the human visual perception of the object itself. The most prominent of these attributes are colour, gloss, translucency and texture. Texture, which is defined as the aspect of appearance that gives rise to a perceivable local variability, plays a complex, yet fundamental, role in human recognition of material characteristics and in pattern discrimination. As a result, its study has gained considerable relevance during the last decades, though it is the most understudied among the visual appearance characteristics. The objective of this Ph.D. thesis is to contribute to the identification of a general texture metric, with particular focus on fabrics.

We first focused on the physical requirements of a texture measurement device. In particular, we initially derived the characteristics required by a multi-spectral system to optimally classify the texture of fabrics. We then showed that the minimum texture classification error can be reached with less than ten spectral channels. Subsequently, we quantified the effect of the radiometric calibration and other standard steps of a camera imaging pipeline on the classification accuracy of trichromatic images. According to the results of the study, the process that most affects the accuracy appeared to be colour rendering.

Afterwards, we analysed the limits and capabilities of the mathematical approaches used to study texture. Specifically, we examined the concept of 'texture stationarity', which, in the context of appearance, is ambiguously defined. We hence proposed and tested a scale-dependent stationarity test for images based on the locally stationary two-dimensional wavelet model. This method was then applied to regular and irregular textures, thus demonstrating that stationarity data can be useful

to improve the accuracy of image classification, especially when classes have been chosen by humans. We also evaluated the relevance of stationarity by correlating it with texture similarity data obtained from a psychophysical experiment.

To conclude, we gathered data on the semantic terms used by fabric experts to describe the visual appearance of textile samples. From these descriptors we then derived a vocabulary of appearance attributes and a fabric ontology. The corresponding descriptors were analysed and linked to a set of texture features, among which CNN ones appeared to perform the best. Finally, we developed a series of simple models based on these results with which it is possible to infer the semantic terms associated with an image.

Sammendrag

Utseendet til et objekt er en kombinasjonen av dets fysiske egenskaper, disse påvirker den menneskelige visuelle oppfatningen av selve objektet. De mest fremtredende av disse egenskapene er farge, glans, gjennomsiktighet og tekstur. Tekstur, som er definert som en måte å betrakte overflate på, gir opphav til en lokal variabilitet. Dette spiller en kompleks, men grunnleggende rolle i menneskelig gjenkjennelse av egenskaper i materialer og i det å skille mønster fra hverandre.

Som et resultat har undersøkelsen fått betydelig relevans i løpet av de siste tiårene, selv om den er den mest undervurderte blant de visuelle utseendeegenskapene. Målet med denne Ph.D. avhandlingen er at den skal gi et bidrag til identifisering av en generell teksturmetrikk, med særlig fokus på tekstiler.

Vi fokuserte først på de fysiske kravene til en teksturmåleenhet. Spesielt avledet vi i utgangspunktet egenskapene som kreves av et multispektralt system for å optimalt kunne klassifisere tekstur av tekstiler. Vi viste da, at den minste teksturklassifiseringsfeilen kan nås med mindre enn ti spektrale kanaler. Deretter kvantifiserte vi effekten av den radiometriske kalibreringen og andre standardtrinn i et kamera-avbildningsprosess på klassifiseringsnøyaktigheten til et trikromatisk kamera. Ifølge resultatene av studien syntes den prosessen som mest påvirker nøyaktigheten å være fargegjengivelse.

Etterpå analyserte vi muligheter og begrensninger til de matematiske tilnærmingene som ble brukt for å studere tekstur. Spesielt undersøkte vi konseptet 'teksturstasjonaritet', som i teksturkontekst er tvetydig definert. Vi har derfor foreslått og testet en skala-avhengig stasjonaritetstest for bilder basert på den lokalt stasjonære todimensjonale wavelet-modellen. Denne metoden ble deretter brukt på

vanlig og uregelmessig tekstur, og demonstrerte dermed at stasjonæritetsdata kan være nyttige for å forbedre nøyaktigheten av bildeklassifisering, spesielt når klasser er valgt av oss mennesker.

Vi evaluerte også relevansen av stasjonæritet ved å korrelere den med teksturlikhetsdata oppnådd med et psykofysisk eksperiment. For å avslutte, samlet vi data om de semantiske begrepene som brukes av stoffeksperter for å beskrive det visuelle utseendet til tekstilprøver. Fra disse deskriptorene hentet vi deretter et ordforråd med utseendeattributter og stofftekstologi. De korresponderende deskriptorene ble analysert og knyttet til et sett med teksturfunksjoner, blant dem syntes CNN å fungere best.

Til slutt har vi basert på disse en serie enkle modeller som det er mulig å utlede de semantiske begrepene knyttet til et bilde.

Acknowledgments

The thesis you are reading is the result of a work that lasted four years, which would not have been possible without the help and support of many people. In this small space I try to thank as many of them as possible.

First of all, I would like to express my gratitude to my supervisors Professor Phil Green and Associate Professor Peter Nussbaum. Thank you for the invaluable feedback, support, and opportunities given to me throughout my study. You have always directed and pushed me to improve, even when I thought progress was impossible. I would also like to include Markus Barbieri in these thanks, as he has been a central figure in the path of this Ph.D. Without his enthusiasm and interest this thesis would never have been written.

Secondly, I would like to express my deep gratitude to those who helped me develop my ideas, taught me to question myself and pushed me to defend the ideas I believed in, such as Hilda, Fereshteh, Helene, JB, Vlado, Noël and many more. I also want to thank all my co-authors for the interesting and exciting collaborations. Thanks to my fellow Ph.D. students and friends at the Colourlab: Ahmed, Mohib, Jacob, Greg, Irina, Tanzima and all the others. Special thanks to Helene and Alf, who put up with me and supported me during my periods in Gjøvik. Moreover, I would like to thank my colleagues and friends in Bressanone, who have made these four years in Südtirol pleasant. Thank you, Andrew, for helping me with the thesis. Furthermore, I am grateful to all the members of Barbieri electronic for their support and help.

Finally, I would like to thank my family and friends for always being present. Thank you, Francesca, for having shared this whole journey with me. Thanks to

mom, dad and Gio for constantly supporting me throughout my entire life. And thanks to the Telegram group.

Contents

Chapter 1

Introduction

The Human Visual System (HVS) provides us with the sense of sight and the ensuing ability to interpret our surrounding environment. This is achieved thanks to the interaction between visible light and the objects present in our field of view, which permits us to infer information from the scene. Within the research field of perception, the mechanisms that make this possible are scientifically referred to as *appearance*, which is formally defined as "the aspect of visual perception by which objects are recognised" [9]. In describing this term, the International Commission on Illumination (CIE) states that through it "an object is perceived to have attributes such as size, shape, colour, texture, gloss, transparency, and opacity" [8]. Moreover, Pointer introduced the concept of *total appearance* [201], a quantity that includes all and only the measurable visual properties, thereby allowing a more complete interpretation of the appearance of an object inserted in a scene.

This thesis has been written as part of an industrial computer science Ph.D. program aimed at defining a way to measure and classify texture, with particular focus on fabrics. This project is the result of the collaboration between the Italian company Barbieri electronic snc/OHG and the Norwegian Colour and Visual Computing Laboratory at the Norges Teknisk-Naturvitenskapelige Universitet (NTNU) in Gjøvik. The present chapter gives a brief introduction to the research work done during the Ph.D. It is subdivided into Motivation (Sec. 1.1), Goals and Research Questions (Sec. 1.2), List of Publications (Sec. 1.3) and Thesis Organization (Sec. 1.4).

1.1 Motivation

Visual perception is one of the five senses with which humans navigate the world. The human eye is sensitive to photons with wavelengths in the visible spectrum domain ($\lambda \in\sim [380, 750]\, nm$). The photoreceptor cells in the retina absorb these photons and convert them into signals that are, in turn, processed by the central nervous system [79], which consists of the optic nerve, the optic tract and the visual cortex. This procedure involves a wide number of tasks, including but not limited to the generation of neural representations, the management of colour vision, the assessment of distance through stereopsis, the recognition of objects and patterns and the perception of motion. All these mechanisms allow us to move in the surrounding environment by inferring its properties and the properties of the objects contained within it through their appearance. Thanks to appearance, humans are extremely skilled at assessing if a surface is near or far, soft or rigid, or smooth or rough. These abilities are part of our daily life and influence our social behaviour and choices as consumers. Their understanding is therefore of utmost importance not only on an academic level but also on an industrial one.

As anticipated above, appearance encompasses various visual attributes. Among these, the CIE identified four main features [78]: colour, gloss, translucency and texture. Usually, appearance attributes are studied within the framework of *soft metrology*. This field aims to develop a metric that links the output of a measurement device (which quantifies a physical stimulus) with the responses of the HVS to the same stimulus [201]. Among the four appearance features, texture, which is linked to the spatial complexity of a scene, is considered the most difficult to address, given that it refers to cognitive processes of a higher level than the other three. This difficulty explains why different definitions of texture have been given in various publications [9, 205, 94, 198, 202], but never in a convincing and standardized way. The physical phenomenon behind texture can be determined by the non-uniformity of either illumination, reflectance, transmittance or topography of the surface under study [102]. This is the reason why the CIE has made a distinction between *surface* [226] and *sub-surface* [223] texture, the former being related to the spatial structure of the sample's reflectance, the latter to its translucency. It is therefore clear that texture is strongly linked to the other appearance features [201]. The mechanism of texture perception is perhaps even more intricate since the significance of spatial cues depends on the context in which they are applied. For example, periodic fabrics can be modeled with spectral-based approaches while natural textures often exhibit fractal structures. For the sake of this thesis, texture is defined as the aspect of appearance which gives rise to a perceivable local vari-

ability [56]. This interpretation should be broad enough to include the previously mentioned definitions.

In literature, texture is mainly used in two different contexts: empirical measurement and computer vision. Empirical texture measurements are employed by the manufacturing industry to evaluate the mechanical properties of a surface [126, 129]. Computer vision, on the other hand, makes use of texture information for various purposes, such as 2D pattern reproduction [198], defect detection [51], classification [115], segmentation [235] and 3D rendering [111]. Texture appearance, however, is underresearched, in part due to its presently unclear definition. Given that the understanding of how it is assessed remains at such an early stage, we have adopted and aimed to provide multiple solid bases for future research in this field. Firstly, to achieve this goal, we decided to focus the attention of this Ph.D. project on fabrics, for which texture is the main appearance trait (as suggested by the shared etymological origin of the words *textile* and *texture*). Secondly, since texture is a high-level aspect of appearance, we have considered it both from a metrological and a computational perspective and we have tried to establish a link between the two.

Additionally, the motivation for choosing fabrics as the study material for this project came from Barbieri electronic. Barbieri is a leading manufacturer of spectrophotometers for the chromatic calibration of medium and large format printers, and has recently dedicated itself to the field of cloth measurement, as the market for digital printing on textiles is at present in great expansion. In this context, Barbieri electronic's aim is the development of a texture metric for the characterization of garments and the optimization of colour reproduction in digital textile printing.

1.2 Goals and research questions

As discussed in the previous section, the main goal of this work is to contribute to identifying a general texture metric for fabrics. We approached this goal by defining the following three sub-objectives:

- *Study the physical requirements of a texture measurement device.*
  There are many computer vision techniques able to quantify texture parameters. These algorithms are usually applied to RGB images but are, however, limited by their capability to reproduce a scene with respect to the HVS. Using digital cameras, for example, the accurate measurement of the tristimulus coordinate of a colour is a challenging task, because they are usually unable to satisfy the Luther-Ives conditions [212]. These require

that the camera sensor sensitivities would be obtainable as a linear combination of the CIE colour matching functions, a precondition not ordinarily satisfied by commercial cameras. The easiest way to solve this issue is to adopt a multi-spectral system. Furthermore, a system aimed at reproducing the HVS should include a method to evaluate the topographical structure of a surface. This poses a problem similar to the one indicated above for the multi-spectral case, but in relation to the height map of the sample. Another possibility to take into account is the choice of geometry of the system because the view of a surface from different angles allows one to estimate the behaviour of the General Reflectance Function (GRF) [102]. Nevertheless, the measurement accuracy is limited by the increased complexity of the system and the cumbersomeness of the measurements.

- *Clarify limits and capabilities of the mathematical approaches used to study texture*.
  Various texture characterization methods have been developed through the years [198]. Different approaches rely on distinct concepts and usually make assumptions about the nature and characteristics of the process that have generated the texture. For example, statistical techniques are generally based on mathematical models, and the model choice is usually linked to prior knowledge of the type of texture to be analysed. These assumptions are, however, not always satisfied and seldom verified. As an alternative, it is possible to analyse texture with learning-based techniques, which have been widely used in the last decade with astounding results in many areas related to computer vision. However, we could not rely on these techniques in this project because they have a strong dependence on the training dataset and because the interpretation of the features they extract is ambiguous.

- *Gather data on textile texture perception*.
  Currently, the biggest obstacle in developing a soft metrology model of texture appearance is the lack of data. Several texture image datasets exist, but only few visual and psychophysical experiments have been performed. It is therefore necessary to collect clues on the response of the HVS to textures with variable characteristics. Once these data are obtained, it is possible to start correlating them with texture features obtained via different measurement and analysis methods.

We can then propose three sets of research questions, each one related to the respective sub-objective:

1. What are the possibilities and limitations of spectral measurements of textile texture? How much do the physical characteristics of the adopted measurement device affect the results of texture analysis? Which texture feature performs best in the classification of textiles?

2. What are the limitations of texture statistical analysis methods? Can they be linked to perceptual stimuli? Can one use the knowledge on these limitations to improve texture classification?

3. Which appearance properties do observers use to describe the texture of textiles? Can texture features be used to model these properties?

1.3 List of publications

This thesis is based on eight articles addressing the research questions discussed in Sec. 1.2. Seven of these have been published through peer-reviewed publication channels, whereas the remaining one is currently under revision at a scientific journal. Each publication is referred to with a letter of the alphabet, and the order follows their occurrence in the thesis narrative. The contributions are divided into core publications, **Articles A** to **E**, whose first author is the writer and which play a central role in the narrative of this thesis, and related publications, **Articles F** to **H**, which correspond to contributions that are part of the work carried out during the Ph.D. and are useful for the discussion, but do not, however, address the problem of texture directly. A rationale to locate these articles in the context provided in Sec. 1.2 and to define the relation between them is illustrated in Fig. 1.1, whose description can be found in Sec. 1.4. The content of the articles is summarized in Chapter 3. Journal articles are shown in boldface, while conference articles are shown in regular typeface.

List of the core contributing articles:

Article A Michele Conni, Helene Midtfjord, Peter Nussbaum, and Phil Green. Dependence of texture classification accuracy on spectral information. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

Article B **Michele Conni, Peter Nussbaum, and Phil Green. The effect of camera calibration on multichannel texture classification. *Journal of Imaging Science and Technology*, 65(1):10503–1, 2021**

Article C Michele Conni and Hilda Deborah. Texture stationarity evaluation with local wavelet spectrum. In *London Imaging Meeting*, volume 2020, pages 24–27. Society for Imaging Science and Technology, 2020

Article D Michele Conni, Hilda Deborah, Peter Nussbaum, and Phil Green. Visual and data stationarity of texture images. *Journal of Electronic Imaging*, 30(4):043001, 2021

Article E Michele Conni, Peter Nussbaum, and Phil Green. Textile texture descriptors. *Manuscript under review in a journal*

List of related articles:

Article F Nadile Nunes de Lima, Michele Conni, Phil Green, and Markus Barbieri. Measurement uncertainty for printed textiles. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

Article G Max Derhak, Phil Green, and Michele Conni. Color appearance processing using iccmax. *Electronic Imaging*, 2018(16):323–1, 2018

Article H Oswald Lanz, Fabian Sottsas, Michele Conni, Marco Boschetti, Erica Nocerino, Fabio Menna, and Fabio Remondino. A versatile multi-camera system for 3d acquisition and modeling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:785–790, 2020

1.4 Thesis organization

This thesis is organized in a compilation of published articles. Fig. 1.1 provides an overview of the research structure and the publications originated from specific topics of interest. This structure has been based on the soft metric framework, which is usually adopted in appearance research [201].

The articles reported in this thesis concern the four steps of the soft metrology framework. In particular, **Articles A** to **E** constitute the core of the work. **Articles A** and **B** attempt to find the optimal spectral setup in relation to the features used for texture classification. They are therefore more focused on the measurement aspect of texture perception, and on the first goal of the project. **Articles C** and **D** provide an analysis of the concept of *texture stationarity*, which until now has been

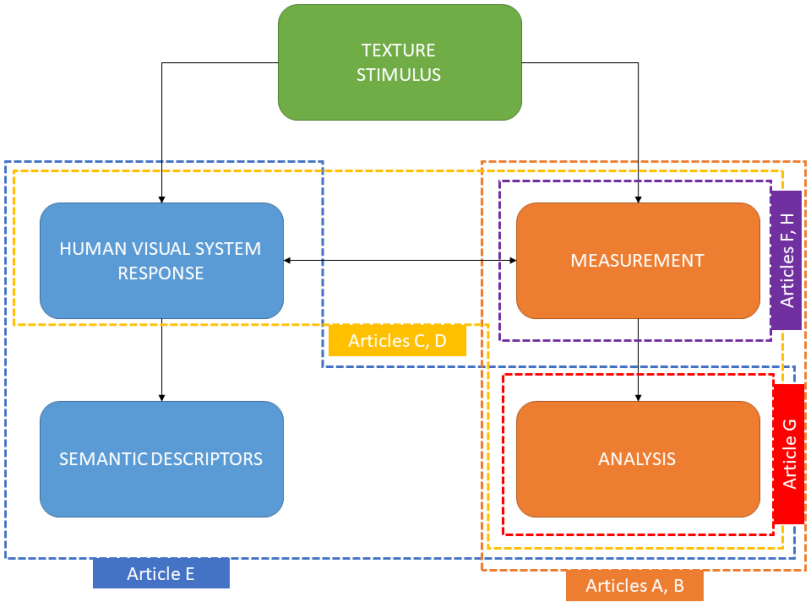


Figure 1.1: An overview of the articles and their relationship to the topics of interest of the research project.

used to indiscriminately describe two separate notions, one mathematical and the other perceptual. In these publications, we proposed to split the two notions and we analysed the correlation between them. Therefore in Fig. 1.1 their scope covers both the perceptual and the analysis fields (although the measurement has not been considered explicitly in the two publications) and addresses the second goal of the thesis. Finally, in **Article E** we studied the vocabulary of fabrics appearance by performing a semantic experiment with experts in the field of textiles as observers. The results of the experiment can be used to estimate the relationship between texture features and semantic descriptors. The remaining articles offer details that are less crucial but nonetheless relevant to the objectives specified in Sec. 1.2. **Article F** is a study on the performance and uncertainty of colour measurements of textile samples. **Article G** analyses different adaptation transforms with the most recent International Color Consortium (ICC) colour management framework, iccMAX. **Article H** describes a novel multi-camera system for the acquisition of the three-dimensional structure of objects. While these do not directly address measurement of texture appearance, they do provide useful background for future research on the topic.

In Chapter 2 we present an overview of the background of the research carried out in the Ph.D. project and of the methods used. Chapter 3 reports a synthesis of the contributions discussed in the previous paragraph, contextualizing them with respect to the research questions and goals discussed in Sec. 1.2. Finally, Chapters 4 and 5 respectively discuss the results of the article and the conclusions of the Ph.D. work.

Chapter 2

Background

In this chapter, we provide an overview of the foundations of literature on which our work rests. We begin with a brief review of what has been done in the context of texture appearance (Sec. 2.1). We then give a short description of the possible typologies of textiles and of their manufacturing processes (Sec. 2.2). Next, we have a summary of the state of the art of texture measurement (Sec. 2.3) and analysis (Sec. 2.4). We then conclude the chapter with the outline of the essential references on which the methods used in this thesis are based (Sec. 2.5).

2.1 Texture appearance

A perceptual stimulus stems from the interaction of the structure of a scene (determined by the spatial arrangement of its atoms and molecules) with its environment (e.g., the illumination). The retinal and neural characteristics of an observer convert the stimulus into an appearance response [118]. The structure and the stimulus of a scene may be accurately measured using hard metrology [201, 37], and the physical properties obtained from these measurements can be used to generate a precise digital reproduction of surfaces and materials [67]. However, these properties do not allow one to predict the human perception of the scene, thus limiting our ability to reproduce the same appearance under different conditions of observation. In fact, we do not know enough about the neural and psychological factors that determine appearance to predict the preference of an observer from the physical characteristics of a structure. To meet this need, Pointer defined the concept of soft metrology as "the measurement of parameters that, either singly or in combination, correlate with attributes of human response" [201]. This discipline

aims at developing measurement techniques and mathematical models that enable objective quantification of the properties of materials, products and activities that are determined by human response (in any of the five senses) [78]. Soft metrology relies mainly on psychophysics, which is "the study of the functions relating the physical measurements of stimuli and the sensations and perceptions the stimuli evoke." [9].

Currently, the most successful and advanced application of soft metrology is in colour appearance. The first attempts to introduce a colour perception model led to the development of colorimetry, which serves as the fundamental underpinning of ensuing approaches. Colorimetry was standardised in 1931 by the CIE with the definition of the CIE XYZ tristimulus space [218]. Since then, more accurate and complete models that take into account finer aspects of colour appearance have been and are being proposed and promoted [79]. With respect to colour, both gloss and translucency appearance models are at a preliminary stage of study, as a satisfactory metric has not yet been found for either of them. Nevertheless, gloss can rely on measurement methods and optical instruments that have been used for decades by many sectors of industry and are still in use today for quality control [157]. These methods have proven to be unable to identify every perceptual cue on which gloss appearance relies, therefore this field is currently the subject of study and discussion [34]. The study of translucency, on the other hand, despite being at a less advanced state than that of gloss, is starting to accumulate a substantial amount of literature [91], including some recent attempts to develop a standard perceptual model [238].

Unlike the fields of gloss and translucency, that of texture appearance can presently rely on a relatively limited bibliography. In the principal publications of soft metrology, texture is usually described as very hard to address [78], and it is approached from the viewpoint of computer vision [201]. Our understanding of perceptual texture in images can be traced back to Bela Julesz, who studied it in the context of texture discrimination. Julesz analysed pre-attentive perception of grey-scale images and tried to understand what makes two textures with the same illumination-related properties distinguishable. He concluded that human texture characterization is correlated to first- and second-order spatial statistics of the image [137, 138], i.e., that humans are not able to distinguish two textures if their first- and second-order statistical parameters are identical. First-order statistics indicate the probability distribution that a dot thrown randomly on an image would fall on a certain grey-scale level, while second-order statistics represent the joint probability distribution of an image, i.e., the likelihood that the vertices

of a segment with a given length and orientation would fall on a certain pair of grey levels. Julesz later noticed that his conjecture was exact in most cases, but with some exceptions. He therefore updated his theory to include the concept of textons [139, 15], which are local geometric structures used by the HVS to discriminate even textures with same second-order statistics. This suggests that a good texture perception model should take into account both global and local cues. The work of Julesz on texture segregation has then been expanded by many other researchers, such as Landy [149], who proposed a model based on non-linear filtering, Rosenholtz [205], who focused on the effect of texture in peripheral vision, and Zhu [254], who linked Markov random field models of texture to the conjectures of Julesz. With respect to texture perception, the main limitation of these approaches is that they are mainly focused on RGB photographs of samples, thus neglecting all 3D cues on which texture appearance is based. In doing so, they neglect the effect of the characteristics and limitations of the measurement device used to acquire the scene, which is a central element of an exhaustive soft metrology model.

Haindl and Filip propose a completely opposite approach to the study of texture appearance by focusing on the precise measurement of the reflectance of a material [102]. To achieve this, they propose a set of physical functions, starting from the General Reflectance Function (GRF), a complex function of 16 variables that describes all possible interactions between the sample and a ray of light striking it. However, the GRF is too complex to be accurately measured or modeled, thus it is usually simplified by applying various simplifying assumptions to the model. The most advanced function used to characterize texture is the Bidirectional Texture Function (BTF) model, a seven-dimensional function that currently can be both measured and modeled. Psychophysical experiments have been applied to BTFs, either to estimate to what extent the dimensionality of a BTF can be reduced to without a noticeable difference in the rendering [81], or to analyse where human gaze fixes when examining a rendered sample obtained with a BTF model [83]. Ref. [82] then links the results of these rendering-based experiments to the same procedures applied to real objects. Moreover, Ref. [64] proposes a soft metrology model based on a variation of the BTF. Despite these simplifications, however, the actual BTF measurement of a material is lengthy and technically difficult to achieve, especially for textiles [32]. This is in stark contrast with how quickly and seemingly effortlessly the HVS assesses the appearance of a textured surface.

In the industry, the main reference for colour management is the ICC. Formed in 1993, the purpose of this organization was to create an open, vendor-neutral colour

management system which would work transparently across all operating systems and software packages [213]. This was achieved through the so-called ICC profiles, files in which the data that characterizes a colour input or output device or a colour space are stored [96]. These files are composed according to standards promulgated by the ICC, and are the final output of radiometric calibration processes. For digital printing, this operation is usually performed with a spectrophotometer, which is the main product of Barbieri electronic. Recently, the ICC developed a novel profile framework, called iccMAX with the aim of expanding colour management to a wider range of applications [62]. This includes, for example, spectral characterization [100] and BRDF information [99]. Yet iccMAX does not consider texture appearance processing, although it permits normal and height map data for 3D rendering to be attached to an image [124]. In **Article G** we contributed to define the capabilities of iccMAX.

As discussed by Pointer in Ref. [201] (see Figure 39 in the publication), an objective assessment of a texture appearance metric can be divided in two processes: the physical measurement of data and the computational analysis applied to them. Therefore, in the following sections we provide a review of both state of the art measurement devices (Sec. 2.3) and computer vision techniques (Sec. 2.4) related to texture. The aim of an appearance metric is to find the mathematical model that permits to infer the HVS response to a stimulus from the output of the measurement analysis. Given that the most prominent of these responses can be expressed through semantic elements, in **Article E** we gathered information on the connection between computer vision features and semantic descriptors by asking a panel of experts to describe a set of physical samples.

2.2 Introduction to fabrics

Formally, one refers to textiles as products formed by the interlacement of fibers, filaments, yarns, woven, knitted and braided cloths as well as non-woven fabrics [186]. Textile materials can be divided in natural-based fibers (e.g. silk, wool, cotton, etc.) and man-made ones (e.g. nylon, polyester, etc.). The textile manufacturing process usually begins with the yarn production, i.e., a long continuous length of interlocked fibers, with or without twist. Once the yarn is ready, it can be used to form a woven fabric through weaving, i.e., the interlacement of two set of yarns perpendicular to each other (the warp and the weft) [2], or knitting, in which the yarn is bent into loops that are then interconnected to form the fabric [221]. Non-woven fabrics, on the other hand, are manufactured directly from the fibers, and are therefore faster and cheaper to produce [207]. These are widely

used, and are mostly made of man-made materials (e.g., polymers), though some of them, for example felt, can be obtained from natural ones, such as wool. From an appearance point of view, while woven and knitted fabrics have a periodic or quasi-periodic structure, non-woven ones are irregular and thus require statistical methods of characterization. Finally, fabrics can be additionally processed in various ways (e.g. singeing, bleaching, printing). The type of material and the post-processing deeply influence the reflectance function of the fabric sample.

Recently, the digital textile printing market has grown considerably [31]. This has led to a push towards improving the accuracy of colour management of textile printers. However, the precise measurement of the colour of fabrics is challenging, because their interactions with light are very complicated [4], and they have a strong influence on the colour coordinates measured by a standard spectrophotometer [169]. We have studied the uncertainty of colour in textiles in **Article F**.

2.3 Texture measurement

As discussed in Sec. 2.1, measurement is the first of the two steps that permit one to derive the objective parameters of a soft metrology model. A texture sample can be measured with many different techniques, depending on the property of interest [154, 102]. Since we are interested in texture elements visible up-close to the naked eye (fabrics are usually observed in such a context), we can set the resolution boundaries of our measurement system between $10\,mm$ and $0.1\,mm$ [37]. Indeed, human vision has a lower resolution limit roughly equal to $0.07\,mm$, whereas surface features smaller than this cannot be directly detected by the unaided eye. Small scale roughness influences the point reflectance function of the sample, and through it its total appearance, by affecting the perception of haze and gloss rather than texture [34].

As clearly exemplified in Adelson and Pentland's *workshop metaphor* [3], the visual stimulus generated by a spatially complex scene depends on three main factors: the spectral and geometrical properties of the light illuminating the scene, the topographical structure of the surfaces contained in it and their reflectances. Since the illumination of a measuring instrument can usually be controlled, in this work we assume that the light source of the scene does not influence the texture pattern, i.e., that it does not project a pattern on the surface under study. However, in practically defining a texture measurement device care must be taken in designing its illumination as it has been proven that texture perception is greatly affect by illumination geometry [36]. Therefore, excluding that, the two physical properties

of a surface that generate a texture stimulus are spatial variations in its topographical structure (*physical texture* in Ref. [201]) or in its spectral reflectance (*optical texture* in Ref. [201]). A review of state of the art measurement of physical texture is reported in Sec. 2.3.1 and of optical texture in Sec. 2.3.2.

2.3.1 Topography measurement

Although the term *binocular vision* literally means *vision with two eyes*, it is usually employed to refer to the limited group of animals who possess a large area of binocular overlap rather than to animals with two eyes [112]. The brain of these animals processes the disparity between the two images and the vergence position of the eyes to infer the distance of visible objects [12]. In addition to depth perception, binocular vision also provides other advantages, both for basic [136] and complex [215] tasks. In this context, textured reflectance information actually provides additional cues to discern the topography of a surface, which is the principle that *shape from texture* algorithms use [5].

Topography measurement methods of textured surfaces have been standardised in ISO 25178, which is comprised of various parts. In particular, part 6 subdivides the available methods of surface texture measurement into three classes [130]:

- **Line profiling**: procedures that have an height function $z(x)$ as output (e.g. stylus instruments, phase-shifting interferometry [58], optical differential profiling [184]). These approaches were the first to be developed [154], but they provide only partial data on the structure of the sample.
- **Area-integrating**: methods that produce numerical results that depend on area-integrated properties of the surface texture (e.g. total integrated scatter [108], angle resolved scatter [243], parallel plate capacitance [155]).
- **Areal topography**: techniques that produce height functions $z(x, y)$ (e.g. parallel stylus measurements [21], coherence scanning interferometry [57], scanning electron microscopy [73]). They are standardised in Ref. [129].

Note that contact-based measurement techniques, such as stylus instruments, cannot be applied to soft and fragile surfaces, such as most fabrics. Given the range of scales of interest, the 3D structure of fabrics has been mainly measured either with stereo systems [142], with photometric stereo [249, 140], or with structured light [1]. In **Article H**, we helped develop a multi-camera system for the reconstruction of the 3D structure of an object, which can be applied to fabric samples.

2.3.2 Reflectance measurements

The reflectance $R(\lambda)$ of a material is defined as following:

$$R(\lambda) = \frac{\Phi_e^r(\lambda)}{\Phi_i^r(\lambda)} \tag{2.1}$$

where Φ_e^r and Φ_i^r are respectively the light flux (or power) reflected by the surface and the flux incident on it [110]. Therefore, $R(\lambda)$ is a function that indicates the effectiveness of the material in reflecting radiant energy. Given a detector with a set of sensitivities $s_n(\lambda)$, the signal acquired by its n^{th} channel can be calculated as [151]:

$$\rho_n = \int_\Lambda E_e(\lambda) s_n(\lambda) d\lambda = \int_\Lambda R(\lambda) E_i(\lambda) s_n(\lambda) d\lambda \tag{2.2}$$

Here, $E_e(\lambda)$ is the spectral irradiance emitted by the sample under analysis, $R(\lambda)$ is its reflectance, $E_i(\lambda)$ is the spectral irradiance of the illumination source and Λ is the spectral domain of interest. The reflectance can be therefore calculated dividing the emitted irradiance $E_e(\lambda)$ by the irradiance of the light source $E_i(\lambda)$ incident on the sample [196]. If the n^{th} spectral sensitivity is the Dirac delta function $\delta(\lambda)$ centered in λ_0, i.e. $s_n(\lambda) = \delta(\lambda - \lambda_0)$, the response of the n^{th} channel of the device corresponds exactly to the spectral irradiance at the same wavelength, i.e. $\rho_n = E_e(\lambda_0)$. Although this theoretical case is not achievable in practice, it shows that the responses ρ_ns of an imaging device provide an estimation of spectral irradiance of a scene, and thus of the radiance of the objects contained in it. Eqn. 2.2 also applies to the HVS by taking the colour matching function (CMFs) $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ of the three cone types as sensitivities s_n [222]. In this case, the wavelength domain Λ corresponds to the visible spectrum, i.e. $\Lambda \in [380, 740]\, nm$ [116].

In the context of texture appearance, as discussed in Sec. 2.3, the spatial variation of spectral reflectance $R(\lambda)$ is one of the two physical properties of the sample that forms the texture stimulus. In fact, a spatially varying reflectance $R(x, y, \lambda)$ can generate a colour texture pattern on the surface (referred to as *optical texture* in Ref. [201]). Therefore, its assessment is at the core of most texture measurement and analysis methods. Currently, the most common way to evaluate the spectral behaviour of the reflectance $R(x, y, \lambda)$ of a surface is with spectral imaging devices, which provide an estimation of the spectral irradiance $E_e(x, y, \lambda)$ of a scene through the model described by Eqn. 2.2 [101]. These were first introduced around the mid 19th century by Pierre Jules César Janssen, a French astronomer who used a double-slit monochromator to analyze the spectrum of the

solar corona [134]. This measurement setup based on point scanning had two major drawbacks: poor light collection efficiency and high sensitivity to motion artifacts. The problems of the monochromator were eventually solved with the introduction of 2D detector arrays in the 1980s [92]. Such devices provided a higher efficiency and the possibility to cover two dimensions of the three dimensions occupied by a spectral image (i.e. the x and y coordinates of the imaging plane and the wavelength dimension λ). Various solutions have been developed to probe the missing dimension:

- **Spatial scanners**: instruments based on this approach use a sequential spectral acquisition, i.e. they use a 2D detector and a diffraction mechanism to gather both spatial and spectral information at the same time. The missing spatial dimension is scanned by moving either the detector or the object. Examples of this approach are pushbroom devices [114, 153].

- **Spectral scanners**: devices similar to spatial scanners, but in this case it is the wavelength dimension that is scanned in time. This is usually done thanks to spectral filters [25, 90, 65].

- **Snapshot setups**: the main drawback of the two previous methods is that they work under the assumption of static scene [233]. The most common choice to acquire dynamic data is to use snapshot cameras, whose integration period are able to capture both spectral and spatial information via a time lapse shorter than the temporal resolution requested. A possible approach is the Integral Field Spectroscopy (IFS), used mainly for astronomy applications. IFS is based on the idea of rearranging the voxels composing the domain of the spectral irradiance $E_e(x, y, \lambda)$ in two dimensions thanks either to mirror [244], fiber [240], or lenslet [48] arrays. Other two alternatives include Computed Tomography Imaging Spectrometry (CTIS) [190, 27], based on wavelengths multiplexing with a transmissive diffraction grating and Multi-spectral Sagnac Interferometer (MSI) [147], which exploits channeled imaging polarimetry, and filter-based methods, such as the tunable echelle imager [11] or the image-replicating imaging spectrometer [107]. But the most common solution to this problem is to sacrifice spatial resolution in favor of spectral resolution by superimposing a colour filter array (CFA) in front of the imaging sensor [13, 19]. This latter approach is particularly effective when a limited spectral resolution is required, as seen in the reproduction of the human visual system response.

Digital cameras based on CFA devices are currently the most widespread imaging devices. Accurate colour reproduction with these instruments requires the exact setting of the illumination and the radiometric calibration of the measurement device [96]. The latter is usually achieved either through the measurement of a target with known reflectance or of the spectral sensitivity functions of the digital sensor [128]. However, the two aforementioned requirements are not always feasible (for example, illumination control is not viable for images containing so-called textures *in the wild* [39]). Given that evidence on multiple fronts shows that the choice of colour space can affect the precision of texture classification [33], we studied the effect of accurate radiometric calibration in **Article B**. In **Article A** we furthermore analysed the dependence of textile classification on the number of spectral channels available.

2.4 Texture characterization

Once the height map $z(x, y)$ and/or the spectral reflectance $R(x, y, \lambda)$ of a source are known, one can manipulate it to extract salient information on the nature of the texture [201]. These significant traits, which reflect some characteristic of the surface, are usually referred to as *texture features* [198]. Features are widely used to perform various computer vision tasks, such as classification [115], segmentation [235] or object recognition [35].

In this section, we present a review of the most used feature extraction techniques available. Sec. 2.4.1 deals with methods applicable to regular textures, Sec. 2.4.2 with those generally used with irregular textures [198]. The features addressed in these two sections are extracted from data with one channel (either grey-scale images or height maps), and their extension to the multichannel case is discussed in Sec. 2.4.3. To conclude, we dedicate Sec. 2.4.4 to the description of the state of the art of *stationarity* analysis of a two-dimensional signal, which is the focus of **Articles C** and **D**.

2.4.1 Regular textures

A texture is regular if it has been generated following a deterministic algorithm [198]. Regular textures can be modelled as consisting of repeated texture elements, sometimes referred to as *texels*. From a synthesis point of view, these patterns can be described with shape grammars, which are a formal way to specify the algorithm used to generate them [229]. A shape grammar is composed by two elements: an elementary pattern, which in our case corresponds to the texel, and a set of placement rules. The textured image is obtained by reproducing the texel according to

the rules [103]. In the cases in which these are repeated recursively, the texture is translationally symmetric, and thus can be described through crystallographic group theory [166]. This mathematical framework can be exploited to detect and describe regular and near-regular patterns images [247].

2.4.2 Irregular textures

Most natural textures are irregular, which means that, unlike regular ones, they cannot be defined through a set of repeatable rules [198]. Depending on the knowledge available surrounding the process that generates the texture and the application, a wide variety of techniques can be used to analyze images containing irregular patterns. In this section we provide a quick summary of the most used techniques.

Grey level co-occurrence matrices

Let us suppose to have an image of dimension $N_x \times N_y$, with N_g grey tone levels and with grey-scale values $g(i,j)$ at pixel position (i,j). Its Grey Level Co-occurrence Matrix (GLCM) $\mathbf{p}(d,\theta)$ is defined as the matrix of dimensions $N_g \times N_g$ whose element (n,m) corresponds to the number of times that two image pixels distant d and at an angle θ from each other assume the grey tone values n and m [198]. Salient parameters can then be extracted from each matrix, the most successful of which have been proposed by Haralick in Ref. [104]. GLCM features have been directly inspired by the first Julesz conjecture [137], and are thus related to second-order statistical properties of the texture. The main drawback of the GLCM features proposed by Haralick is that they are very time-consuming to calculate [181]. Therefore, many alternative methods based on the same principles have been proposed through the years [208, 182].

Histograms

Another broadly used group of texture analysis methods represents images through parameters of their histograms [209, 214, 210]. The most popular of these approaches is the Histogram of Oriented Gradients (HOG), introduced in a 1986 patent application [176] and further developed in Ref. [53]. Widely used in the field of object detection, this technique is based on the idea that texture can be described by local histograms of image gradient orientations in a dense grid. As a consequence, however, it is limited in the detection of global structures.

Markov random fields

A Markov Random Field (MRF) is a stochastic process that can describe the spatial relationships between pixels in a user-defined neighborhood [115]. The MRF

model is based on the assumption that the value of a pixel directly depends only on the intensities of its neighboring pixels, or the *Markovian property* [161]. MRFs were first applied to the field of texture characterization by Ref. [50], but they became popular with Ref. [17], in which they were used to model generic lattice systems. Currently, they are mainly employed in classification and segmentation tasks [173]. MRFs characterize an image with a parametric distribution that accounts for the relations with neighboring elements as terms of its prior probability. Since they can only describe relationships between pixels within the neighborhood, the main drawback of MRF-based techniques is that they are suited to a limited number of textures [255].

Fractals

A fractal is an object with irregular structure that shows a degree of self-similarity at different scales [143]. In texture analysis, non-deterministic fractal processes are fitted to the image to extract self-similar statistical parameters, such as the fractal dimension D [135]. Intuitively, D links the number $N(\epsilon)$ of basic elements (e.g. boxes, spheres) required to cover a set $S \in \mathbb{R}$ to their characteristic length ϵ, i.e. $D = log_{\epsilon} N$. In the case of a grey-scale image, this set S corresponds to the bidimensional intensity function $I(x, y)$ and the fractal dimension D estimates how self-similar the image is at different scales [198]. The main flaw of fractal methods is the same as with MRF: they work well only on images that have an irregular geometry, such as those generated with natural processes [115].

Local binary patterns

The Local Binary Pattern (LBP) texture features were first introduced in Ref. [189] and became popular with Ref. [188]. To define the LBP, one needs first to define a standard neighbourhood of P pixels with respect to a central one. Then, each neighbouring pixel is compared to the central and the pattern is transformed in a binary series according to the equation:

$$LBP_{P,R} = \sum_{p=0}^{P-1} \text{sign}(g_p - g_c) 2^p \tag{2.3}$$

Here, R represents the radius of the neighbourhood, $\text{sign}(z) = 1 \Leftrightarrow z \geq 0$, otherwise $\text{sign}(z) = 0$, and g_c and g_p indicate the grey levels for the central and the p^{th} neighboring pixel, respectively. This operator generates a map with 2^P levels, which characterizes each pixel in function of its relative relation with its surroundings. As it describes the distribution of local patterns, the LBP method

can be considered as a texton-based approach [139]. LBP features have proven to be very effective in texture classification tasks, and many approaches have been developed based on these [200, 24]. Their main drawback is their low resilience to noise [251].

Fourier transform

A wide group of texture characterization techniques are based on spectral methods [198]. The forerunner of these methods is the Fourier Transform (FT) operator, which can be defined as:

$$FT[f(x)] = \hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x\xi}dx \tag{2.4}$$

where ξ is the variable of the transform space and i is the imaginary unit [94]. The FT permits to project a function $f(x)$, $x \in \mathbb{R}$ on a complete set of orthonormal elements, represented by the $e^{-2\pi i x\xi}$ term in the definition. This corresponds to decomposing an image into a summation of sine and cosine functions with different phases and frequencies [115]. The resulting function $\hat{f}(\xi)$, $\xi \in \mathbb{C}$, is a function in the frequency domain. The FT is therefore used to represent the frequencies that form an image, which is particularly suitable to describe repetitive and periodic textures [253]. The main defect of the FT in the context of image analysis is that, since it is a sum of sine and cosine functions, it is not well-suited to capture local transient signals.

Wavelet transform

The basic idea behind the wavelet transform is to perform a spectral analysis based on the same mathematical principles of the FT, but with a local functional basis [54]. We can thus define the wavelet transform as as:

$$W_\psi[f(x)] = W_\psi(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi^*(\frac{x-b}{a})f(x)dx \tag{2.5}$$

where ψ is an $L^2(\mathbb{R})$ function, called *mother wavelet*. The mother wavelet can be used to define a complete family of functions $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ with $j,k \in \mathbb{Z}$, which correspond to its translated and stretched versions. $\psi(x)$, and consequently also the functions derived from it, is by definition a compact support function with oscillatory characteristics, and it is therefore spatially limited [185]. Generally, the mother wavelet is associated to a low-pass filter, and it is paired with an auxiliary function $\phi(x)$ called *father wavelet*, which is added to the wavelet family to obtain a complete functional basis [237].

Gabor filters

Gabor filters are a group of functions based on a complex sinusoidal function modulated by a bidimensional Gaussian distribution [198]. They have been explicitly modeled on early processing stages of the HVS [191]. The basic structure of a Gabor function is the following:

$$h(x,y) = g(x,y)e^{i2\pi[Ux+Vy]} \tag{2.6}$$

Here, $g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\{-\frac{1}{2}[(\frac{x}{\sigma_x})^2+(\frac{y}{\sigma_y})^2]\}$ is a Gaussian function, while the vector $(U,V) \in \mathbb{R}^2$ represents a specific 2D frequency. Mathematically, this family of functions is particularly important because it permits to minimize the wavelet transformation uncertainty principles (i.e. $\Delta x \cdot \Delta u \geq 4\pi$ and $\Delta y \cdot \Delta v \geq 4\pi$) in two dimensions [55]. Therefore, it provides the best trade-off between spatial and spectral localization achievable with a spectral transform. Several authors have modified and improved this texture analysis technique [41, 172, 70].

Autocorrelation

The autocorrelation function quantifies the similarity of a texture with its shifted self as a function of the shift distance [94]. It is able to detect repetitive patterns, and therefore it can be linked to the FT. The shape of the autocorrelation function is used as an indicator of how coarse a texture is [115]. The normalized autocorrelation of an image with dimensions $N_x \times N_y$ can be defined as:

$$\rho(x,y) = \frac{\frac{1}{(N_x-|x|)(N_y-|y|)}\sum_{i,j} I(i,j)I(i+x,j+y)}{\frac{1}{N_xN_y}\sum_{i,j} I(i,j)^2} \tag{2.7}$$

where $I(i,j)$ indicates the grey level of the image at position (i,j). As the FT, the autocorrelation is based on a global model, and it thus tends to neglect local information.

Mathematical morphology

Mathematical Morphology (MM) is a theory used to analyse spatial structures [220]. It can be used to extract information on the shape and form of objects, and it is based on a mix of set theory, integral geometry, and lattice algebra. It has also proven to be a powerful image analysis technique, and can thus be found at the core of many image processing and analysis algorithms [250, 165, 22]. Moreover, MM provides a set of tools, such as granulometry and the pattern spectrum, that can be employed to characterize textures [219]. These quantities supply us with

information on the size of objects in the image by applying the morphological transformations of opening and closing to it. The advantage of MM features is that they can be computed very efficiently [60], while their main disadvantage is that they are unable to store positional information.

Deep learning features

The complexity of texture appearance has prompted many researchers to seek a solution to texture-related challenges in the increasingly widespread Convolutional Neural Networks (CNNs) [164]. A CNN is a deep neural network whose connectivity pattern between neurons is inspired by the organization of the animal visual cortex [175]. Perhaps thanks to this similarity with natural structures, this typology of networks has proven to be very effective in many computer vision tasks [146, 26, 39, 40, 7, 163, 252, 52]. In fact, the authors of Ref. [14] proved that using off-the-shelf features extracted by pre-trained CNNs for texture classification provides, in most cases, better results than handmade ones. The main drawback of CNNs is their black-box nature, which makes it difficult to clearly interpret how the inputs of the network determine its decisions [245]. Given the current state of the art, it is therefore impractical to use these techniques for the development of a texture appearance model.

2.4.3 Multi-spectral features

In Subsec. 2.4.2 we limited ourselves to addressing texture features in the context of grey-scale images. However, it has been proven that taking into account colour information improves the performance of various computer vision algorithms [68, 194, 180]. In fact, most texture analysis methods discussed have been extended to multichannel applications. It is therefore useful to give a summary of the methods used to expand texture features to cases of images with three or more channels.

Ref. [33] divides these techniques into two main categories. The first one is composed by the integrative approaches, which derive intensity features of each channel and join the results into a feature vector. The second one encompasses parallel approaches, which first convert the image to grey-scale, calculating its intensity features, and then merges them with a set of global colour parameters. Pure colour features used in parallel techniques are generally elementary properties, such as descriptive statistics or percentile information. They are applied to the dimensions that contain colour information (e.g. a and b in the CIELab space), which depend on the choice of colour space [194]. Integrative approaches, on the other hand, are directly derived from their corresponding intensity methods [156, 174, 10]. For

both of these categories, the colour space and coordinate normalization methods adopted can have a big influence on the performance of the implemented task. In **Articles A** and **B**, we linked this influence to the choice of a texture measurement setup.

2.4.4 Texture stationarity

Stationarity is a property of signals that represents the independence of the distribution of their generating process from time or space [121]. It is central for texture interpretation because it indicates if and when global methods are applicable [198]. Therefore, many texture feature extraction techniques assume that the image to which they are applied is (second-order) stationary [94]. This is due to the fact that the stationarity of an image informally means that its statistical properties do not change from pixel to pixel [185]. Formally, this translates to a property type called *strong stationarity*, which requires that a signal has a finite-dimensional distribution that is shift-invariant, i.e. that its joint distribution F in a sub-region of the image is the same everywhere ($F(X_{\mathbf{r}_1}, \cdots X_{\mathbf{r}_n}) = F(X_{\mathbf{r}_1+\Delta r}, \cdots X_{\mathbf{r}_n+\Delta r}) \, \forall \mathbf{r}_i, i, \Delta r$). However, strong stationarity is usually too strict to be applied in practice, therefore the concept of weak stationarity is generally used. A signal is weakly stationary if the first and second moments of its process are invariant under a coordiante shift, i.e. $cov(X_{\mathbf{r}_1}, X_{\mathbf{r}_2}) = \gamma(\mathbf{r}_1 - \mathbf{r}_2)$. This is the type of stationarity generally assumed by texture analysis methods.

The stationarity of a signal can be verified with a statistical test [231]. Many stationarity tests have been developed over the years for one-dimensional signals, but only a few of them have been extended to two dimensions. Of these, the approach proposed in Ref. [75] is based on the fact that the second-order cumulant spectrum matrix of a stationary image should theoretically be diagonal. Therefore, the authors used the l_2 norm of its off-diagonal elements as test statistic. This method has been put in practice in Ref. [76], which applied it to the problem of source detection in shallow water. Ref. [23], on the other hand, proposed to evaluate the spatial stationarity of uniform linear arrays using the generalized likelihood ratio statistic. The main drawback of these two methods is that they require multiple realizations of the process to evaluate its stationarity, which is a strongly restricting requirement. The author of Ref. [87] tested stationarity on the basis of the homogeneity of a set of spatial spectra evaluated on different windows of the image. On the other hand, Ref. [20] used the empirical mean and variance of images with known underlying asymptotic behavior to detect any anomalies, which should hint at the images' non-stationarity. The main defect of these latter two approaches is that

they cannot be automated. In fact, the output of the first algorithm depends on how the windows are chosen, while the second one relies on visual inspection to detect the non-stationarity of the images. Finally, Ref. [231] introduces a stationarity test based on the locally stationary two-dimensional wavelet (LS2W) model, using the constancy of its power spectrum as test statistic. According to Ref. [78], an important side of the texture's appearance is its dependence on the viewing scale. Given that the LS2W test is able to assess the stationarity of a single image at different scales, we contributed to the definition of a soft metrology scheme for texture appearance by using it to address the uniformity and scalability of a texture in **Articles C** and **D**.

2.5 Research methods

Each step of the research discussed in this Ph.D. thesis required a specific methodology. In most cases, the adopted methods were conceived as variations and continuations of approaches used by other researchers, whose work was perceived as particularly relevant for the purposes of the Ph.D. project. This approach allowed, indeed, a great consistency with literature as well as methodological rigour.

For what concerns the first part of the project, which addresses the problem of outlining the physical requirements of a texture measurement device (see Sec. 1.2), we used the framework of texture classification to evaluate the efficiency of the measurement system defined. In particular, in both **Article A** and **Article B** we measured a set of hyperspectral images, derived from them the reflectance of the scene and applied to it various spectral sensitivity functions to simulate the response of different imaging systems. This procedure is in line with the methods commonly used to characterize digital imaging devices [213]. The dataset employed in these two papers was obtained with an HySpex VNIR-1800 camera, for which we adopted the measurement approach presented in Ref. [144]. Furthermore, **Article B** and **Article H** assessed the calibration of an RGB camera, which has been performed following Ref. [128]. In the same publication we also implemented a manual rendering and a sharpening through a basic unsharp masking process, as suggested in Ref. [94]. The uncertainty measurements made in **Article F** have been mainly rest on Ref. [131].

Moreover, in these first two articles of the Ph.D., as well as in **Article E**, we extracted a wide variety of texture features from the images obtained. These have been implemented on the basis of various references, in particular Ref. [33], to which the articles are inspired, and Ref. [197], which provides a comprehensive summary of the techniques of texture analysis in computer vision. A survey of

the techniques used can be found in Sec. 2.4.2. The choice of classifiers, 3-NN and random forest, and the colour texture examination method were also mainly based on Ref. [33]. We also employed texture classification in **Article D**, for which however we drew inspiration from Ref. [14]. On the other hand, the computations carried out in **Article G** have all been addressed in Ref. [63].

Articles C and **D** addressed the problem of the evaluation of texture stationarity. To achieve this, we implemented the image stationarity test introduced by Ref [231], which has its foundations on the wavelet research of Nason and colleagues [185, 72]. In **Article C** we also proposed a scale-dependent alteration of the method, which we evaluated according to the criteria used in Ref. [231], as well as tested it on the Amsterdam Library of Textures (ALOT) database.

We performed various visual experiments during the Ph.D. project discussed in this thesis. In particular, in **Article D** we calculated the Spearman rank-correlation between the confusion matrices of the classification algorithms and a similarity coefficient derived from a texture grouping experiment. The results of this process, that we conceived ourselves, were then analysed to quantify the relationship between perceptual and mathematical stationarity. On the other hand, in **Article E** we have collected a series of semantic data on the appearance of fabrics, which we have arranged in a vocabulary on the basis of Ref. [118]. In the same paper, we proposed an ontology for textiles, which was based on the Semantic Web framework [211].

Chapter 3

Summary of the included articles

In this chapter, we summarize the articles included in this thesis. These include three journal papers, five conference proceedings and a series of oral contributions. For each paper, we present a synopsis of motivations, methods and outcomes. For further details, refer to the respective manuscripts, which can be found in the final part of this thesis.

3.1 Core contributions

In this section, we gather the major contributions to the topic of this Ph.D. project.

3.1.1 Article A: Dependence of texture classification accuracy on spectral information

> Michele Conni, Helene Midtfjord, Peter Nussbaum, and Phil Green. Dependence of texture classification accuracy on spectral information. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

As addressed in Subsec. 2.4.3, there is evidence that one can improve the performance of texture analysis by taking into account colour cues. To our knowledge, however, this principle has only been applied to trichromatic images. Given the growing interest in the study of texture of hyper-spectral images, the purpose of this paper was therefore to quantify the value of spectral information in texture classification. Since the dataset used for this investigation was mainly composed of textiles, this analysis also provided an insight on the requirements of a measure-

ment system aimed at the classification of fabrics.

As a database of hyper-spectral texture images was not available, we first acquired a set of 191 hyper-spectral images using a HySpex VNIR-1800 camera [122], some of which are shown in Fig. 3.1. For each image we then extracted the spectral reflectance information of the samples using a reference surface with a known reflectance. A set of simulated multi-spectral images with an increasing number of channels was subsequently obtained from each sample reflectance by applying various spectral sensitivities to them. These sensitivities were defined as the collection of n identical Gaussian functions whose variances covered the range of wavelengths in which the HySpex is sensible. After this, we extracted a selection of texture feature vectors for each simulated image based on various techniques (GLCM, LBP, Gabor and wavelet) with an integrative approach. We eventually classified these vectors with two different classifiers (3-nearest neighbours (3-NN) and random forest (RF)) and collected the accuracy of the algorithm depending on the number of spectral channels. This process was performed both in the whole spectral domain of the camera ($\lambda \in [400,\ 1000]\, nm$), which includes a large portion of in the near infrared (NIR) region, and in the visible ($\lambda \in [400,\ 730]\, nm$).

The results of the experiment show that texture classification accuracy increases with the number of spectral channels with which the image is acquired. This increment has however a saturation value, which depends on the type of feature used. For the features considered, the saturation is reached with less than ten channels. These results are valid for both classifiers, but the 3-NN appears to outperform RF for LBP features. Considering only the visible spectrum on average enhances the classification performance of 5% for the lower number of channels, although the asymptotic value of the curves appears to be independent from it. In order to infer the optimal number of channels that a measurement system should have to maximise texture classification, however, further study is needed, because in this work we used ideal spectral sensitivities. The ideas discussed in this paper have been expanded in **Article B** (Sec. 3.1.2).

The Ph.D. candidate conceived the research idea, gathered the experimental data and proposed the methodology of analysis. He also contributed significantly to the practical implementation of the computations needed for the analysis and authored the article.

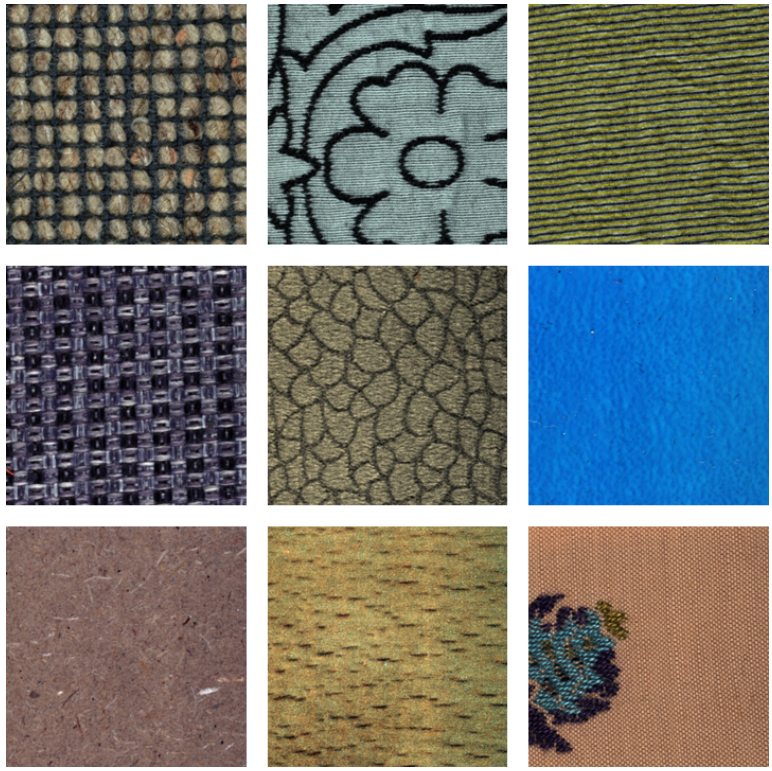


Figure 3.1: Example of the dataset acquired in **Article A**.

3.1.2 Article B: The effect of camera calibration on multichannel texture classification

Michele Conni, Peter Nussbaum, and Phil Green. The effect of camera calibration on multichannel texture classification. *Journal of Imaging Science and Technology*, 65(1):10503–1, 2021

To capture images in a perceptually meaningful colour space, an accurate radiometric calibration of the employed camera is needed [96]. However, this is a cumbersome and lengthy process, and in some situations it may be impracticable. It has been shown that the colour space in which the extraction of texture features is performed can affect the results of texture analysis algorithms. Therefore in this paper we aimed at quantifying the effect of accurate camera calibration and of other standard steps in a camera imaging pipeline on texture classification.

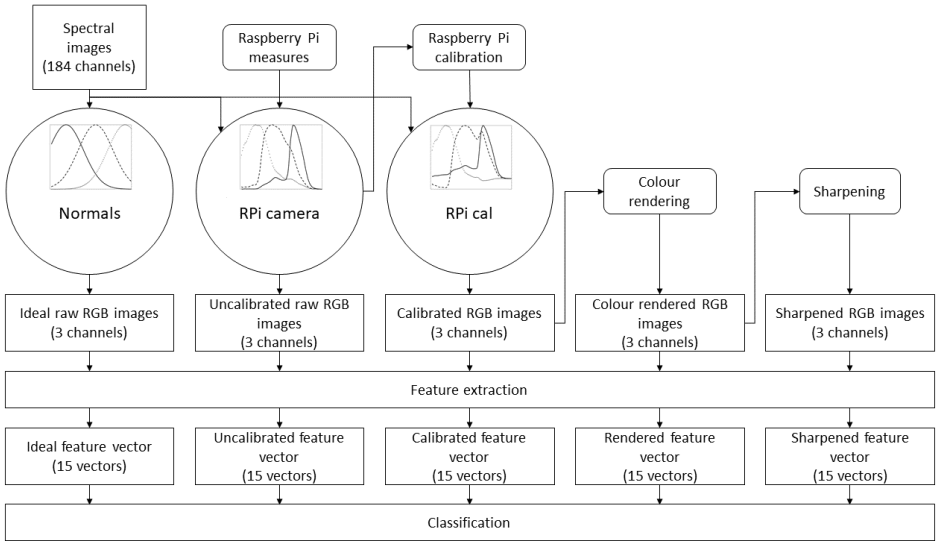


Figure 3.2: A scheme of the experimental workflow followed in **Article A**. The number of elements constituting the data is reported at each step.

This may affect how texture image databases are used, some of which lack calibration data. In this work we also compared the ideal results of **Article A** (Sec. 3.1.1) with those obtained with spectral sensitivities of real devices.

In order to implement this analysis, we first determined the spectral sensitivities of a Raspberry Pi V2 camera module with a monochromator. We then applied them to the database of reflectances acquired in **Article A**, thus extracting a set of uncalibrated trichromatic images. For each of these we calculated the same set of texture features used in **Article A**, which we used to classify the samples with two classifiers (3-NN and random forest). We compared the classification accuracy with the results of **Article A**. Following that, we examined the effect of various standard processes normally applied to images in a commercial camera. First, we optimized the colour matching functions of the Raspberry Pi camera, thus obtaining the model of a calibrated device. We then applied colour rendering and sharpening to the images, classifying them after each step. A scheme of the whole process is displayed in Fig. 3.2.

The results of the experiment show that the accuracies obtained with the measured Raspberry Pi camera sensitivities are very similar to those derived in **Article A**. Moreover, the various steps of the image acquisition pipeline had a relatively

small effect on the classification results. Between these, the rendering step is the procedure that most affects the accuracy, increasing it by approximately 5%. This could be due to the fact that for rendered images the colours are usually more saturated, which could increase the distance between different classes in the features space. These results suggest that, when classifying texture images, it is preferable to work with fully rendered pictures rather than pure raw images, which managed to match the accuracy results achieved by the multi-spectral systems simulated in **Article A**.

The candidate contributed to this paper by proposing the research idea, developing the research methodology and performing the Raspberry Pi camera calibration procedure, both experimentally and computationally. Moreover, he wrote the research paper and followed its process of submission and peer-review.

3.1.3 Article C: Texture stationarity evaluation with local wavelet spectrum

> Michele Conni and Hilda Deborah. Texture stationarity evaluation with local wavelet spectrum. In *London Imaging Meeting*, volume 2020, pages 24–27. Society for Imaging Science and Technology, 2020

Ref. [197] defines a *stationary texture image* as an image that contains only one type of texture. Based on this definition, the authors recommend analysing non-stationary images with methods different from those used for stationary ones. However, this line of thinking is ambiguous, because stationarity can have two possible interpretations. If the authors of Ref. [197] intended stationarity in a mathematical context (see Subsec. 2.4.4), they should have indicated a method for quantifying the stationarity of images. Otherwise, if they meant it in the context of visual perception, they should have referred to the texture discrimination literature (see Sec. 2.1), which, however, does not currently indicate a comprehensive model to evaluate the homogeneity of texture. Given that, as far as we know, the mathematical and perceptual frameworks have never been compared - though they are logically akin - in **Article C** we performed a pilot study in order to probe if and how they are related.

A discrete time series $X_t \in \mathbb{Z}$ is defined as *weakly stationary* if the mean of its joint distribution is constant over time and its covariance between two time coordinates t_1 and t_2 depends only on their difference ($K_{XX}(t_1, t_2) = K_{XX}(\tau, 0), \tau = t_2 - t_1, \forall t_1, t_2 \in \mathbb{N}$). Ref. [72] expanded this concept to image processing by defining it for a bidimensional signal with pixel coordinates (u, v), and Ref. [231] further used it to delineate a stationarity test for images. In **Article C**, we first pro-

posed and implemented a modification to this test to make it capable of extracting scale-related stationarity information. We then applied this test to two groups of 100 artificial images obtained with homogeneous statistical distributions, a white noise process and a correlated Gaussian random field, to confirm that they were actually labeled as stationary. Finally, we applied it to a subset of the ALOT texture database [28], in order to personally judge the effects of various types of texture on the output p-values.

The results of the experiment with the artificial images show a p-value higher than the chosen significance level of 5% at every scale, thus confirming its effectiveness. Furthermore, the $\mathbf{p}$ vector has revealed to be an interesting tool to fathom these attributes at different scales. Additional results of this analysis can be found at this link. From these results it seems that man-made textures are more stationary than natural ones. Based on this study, we planned the experiments performed in **Article D** (Sec. 3.1.4).

The thesis author contributed to the conception of the research idea and proposed the modifications of the stationarity test to add scale dependence. Together with the co-author of **Article C**, he designed the experiment, developed the research methodology, acquired the data and performed the analysis. The candidate also contributed substantially to writing the research paper.

3.1.4 Article D: Visual and data stationarity of texture images

Michele Conni, Hilda Deborah, Peter Nussbaum, and Phil Green. Visual and data stationarity of texture images. *Journal of Electronic Imaging*, 30(4):043001, 2021

The objective of this paper was to further gain information on the LS2W stationarity test proposed in **Article C** (Sec. 3.1.3) and on its relationship with perceived stationarity. This link has been discussed in Ref. [231] in the context of fabric pillage, whose results were neither general nor supported by experimental data. Moreover, in the framework of soft metrology, **Article D** aimed at advancing the development of a model for texture uniformity.

To achieve these goals, we performed three different investigations: a chessboard stationarity analysis, two classification tasks, and a psychophysical experiment. For the first one, we generated a set of chessboard images of various dimensions, orientations and noise levels. With this, we wanted to study in detail which effect a plain regular texture such as a chessboard would have had on the p-values at

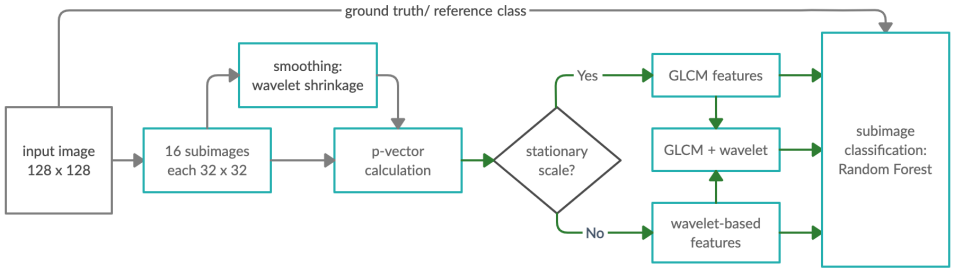


Figure 3.3: Structure of the first task in the classification investigation of **Article D**. Each input image is split into sub-images, which are then classified. The precision of the classification is evaluated based on whether the sub-images are correctly linked to their original image.

different scales. The second analysis was aimed at understanding the veracity and the limits of the assertion that some texture analysis techniques are more suited for stationary image characterization than others. We thus performed two different classification tasks on 110 texture images extracted from the Describable Textures Dataset (DTD) [69, 39]. In the first of these classification tasks, two texture feature techniques, GLCM and wavelet, were used to extract descriptive vectors from each image. In fact, according to Ref. [197], GLCM texture features are more suited to describe stationary images than wavelet. We then compared the classification accuracy obtained considering the stationarity information provided by the LS2W test with the accuracy supplied by a standard classification approach (see Fig. 3.3). We subdivided the original images into non-overlapping sub-images and we used the original image as class. In the second task, we repeated the experiment, but we used the texture classes defined in Ref. [39] and five feature vectors (LBP, GLCM, HOG, Gabor and wavelet). Again, the performances of the algorithm with and without stationarity cues were compared. As a reference, we also performed the same task with a set of features obtained from various pre-trained CNNs. Finally, in the third task we set up an online psychophysical experiment based on the definition of perceptual stationarity given in Ref. [198]. In this we presented to the observers a set of 25 images and a reference, all randomly selected from the subset of the DTD database used for the classification experiment. We then asked the observers to select all the images that looked similar to the reference. From this, we calculated the correlation between the texture similarity judged by the observers and the results of the classification tasks.

The application of the stationarity test to the chessboard images showed a strong

stationarity, which is in line with how the HVS perceives regular textures. However, the $\mathbf{p}$ vector of an image obtained by stretching the chessboard had an unexpected non-stationary artefact at a single scale, which suggested that the test could be improved. The classification experiments provided us with additional insights on the characterization of irregular textures. In particular, the first classification task demonstrated that the approach to texture analysis in stationary and non-stationary images proposed in Ref. [198] has limited validity in a classification framework. On the other hand, the results of the second task suggest that using the stationarity information does improve the classification of the DTD images in their human-defined macrogroups. Finally, although the traditional texture features appear to be weakly correlated with the results of the psychophysical experiment, the technique that achieved the best correlation with the response of the average observer was a mix of wavelet and LBP features obtained with the stationarity information.

In **Article D**, the candidate played a main role in the conception of the research idea, the design of the experiment and the development of the research methodology. He also worked on the algorithms for the analysis of the chessboard images and for the classification tasks, set up the psychophysical experiment, implemented the analysis of data and wrote the research paper based on the results. Finally, he served as corresponding author in the process of submission of the paper at the Journal of Electronic Imaging and worked with the co-authors to its revision in the peer-review process.

3.1.5 Article E: Textile texture descriptors

Michele Conni, Peter Nussbaum, and Phil Green. Textile texture descriptors. *Manuscript under review in a journal*

Fabrics have played a fundamental part in human society for tens of thousands of years. However, the link between physical properties and texture appearance of a textile sample is still far from being fully understood. A perceptual model of fabrics would be extremely useful for appearance reproduction, and could benefit industries active for example in the online shopping and digital printing sectors. In order to obtain information on the behaviour of texture perception, in this article we studied the relationship between optical characteristics of textiles and semantic descriptors used by experts in fabric design to define them. We organized these descriptors in a vocabulary of fundamental terms, and we proposed an ontology in order to build a general structure of textile perception, logically linking it to

physical properties.

The first step of this analysis consisted in gathering textile descriptors. We asked a number of researchers in fabric design to participate in a visual experiment, ten of whom agreed to collaborate. Some of the participants also invited collaborators to join, eventually amounting to 28 participants. We then sent to each group of observers a set of 22 fabric samples, selected between the most commonly used materials and weavings. The observers were asked to identify no more than five English words that they would have used to describe each fabric sample to someone who was not able to see it. Once we received the results from each group, we brought them together, filtered those relative to appearance and arranged them into a vocabulary formed by their macro-groups, which highlighted pairs of descriptors with opposite meaning. Following this, we proposed an ontology for textile appearance, based on objective information related to materials and manufacturing techniques. The purpose of the ontology was to clarify the scope of the terms used by the observers and the logical relation between them. After that, we compared the frequency with which terms corresponding to each macro-group were used to describe the different fabric samples with the value of some popular texture features. We did this by calculating the Spearman rank-correlation coefficient for each feature-descriptor couple, thus obtaining the feature with the highest correlation for each descriptor. These features were extracted from data obtained with two different measurement devices: a simple RGB camera and xTex by Vizoo [246], which acquires various maps of the sample with which one can generate a rendered version of the surface. Finally, we used the results of the analysis to infer the most suitable terms for a subset of the textile image database proposed in Ref. [140].

The results of the psychophysical experiment showed that the observers tended to use not only visual attributes, but also other types of adjectives, which can be related to mechanical and material properties of the samples. The textile ontology, based on fabrics literature and industrial standards, was proposed to uncover the rationale behind these descriptors and to clarify the boundaries between objective (which one can therefore study independently of perception) and subjective (which can be analysed through soft metrology) properties of textiles. On the other hand, the conclusions of the comparison between frequency of use of each descriptor and value of various texture features showed how pre-trained CNN-based features achieve the highest Spearman correlation value. However, this result is linked to the fact that each CNN extracted a total of 1000 features, a much higher number than that obtained with traditional techniques. If we neglect the learning methods, the best correlation was achieved by higher statistical moments of the distribution

of pixel grey-scale values. Moreover, by comparing the results obtained with the two measurement techniques we can conclude that a good texture analysis setup requires not only 2D spatial information, but also topographical data. Finally, the results of the application of prediction models to the textile image database seemed to be reasonable.

This article was conceived and designed by the Ph.D. candidate, who also proposed the research methodology, followed the organization and realization of the visual experiment, envisioned and performed the analysis of data, and wrote the article. The candidate has served as corresponding author in the process of submission of the paper, which is currently under review. We expect to be able to provide a revised version of the article before the Ph.D. defense.

3.2 Minor contributions

In this section we provide a summary of the minor contributions resulting from various collaborations during this Ph.D. project.

3.2.1 Article F: Measurement uncertainty for printed textiles

> Nadile Nunes de Lima, Michele Conni, Phil Green, and Markus Barbieri. Measurement uncertainty for printed textiles. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

Colorimetric measurements of materials with a complex topographical structure, such as wood or fabric, usually employ spectrophotometers with a large sampling aperture, although graphic industry standards do not provide any specific recommendation on the matter [127]. In fact, since the device measures the average colour in its field of view, a larger aperture should correspond to a more stable and reliable estimation. Moreover, the $0° : d$ geometry is usually considered more suited to the measurement of anisotropic samples than the $0° : 45°$ or $45° : 0°$ ones due to their supposed independence from the positioning on the surface under study. In practice, this should correspond to higher reproducibility and lower uncertainty of the measured data. The aim of this paper is to provide experimental evidence to these hypotheses in the specific case of digitally printed fabrics.

We addressed this problem by evaluating the performance of four commercial spectrophotometers (one bench-top and three hand-held) in terms of short-term repeatability, inter-model agreement and reproducibility. To assess the short-term repeatability, we measured various samples: a white Spectralon tile, made of polytetrafluoroethylene (PTFE), a white paper sample with optical brightening agents,

and five raw fabric samples. We then estimated the Mean Colour Difference from the Mean (MCDM) of the data [16]. After this, we calculated the inter-model agreement between all instruments, thus quantifying how accurately two spectrophotometers can repeat the same colour measurement. Finally, we evaluated the reproducibility of each device by changing the measurement position on the five textile samples and calculating the MCDM.

On the PTFE tile and the white paper sample all instruments performed consistently with the manufactures specifications, while on the textile samples the short-term repeatability was in most cases of the same order as on the Spectralon tile. Varying the position of the sample with respect to the instrument aperture resulted in a higher uncertainty, which depended on both the device and the textile material. On printed samples, the $0^\circ : d$ instrument performed best in terms of short-term repeatability, regardless of the orientation to the sample. However, in terms of reproducibility, i.e. when measuring the same sample at different locations, the diffuse spectrophotometer showed similar results as $45^\circ : 0^\circ$ ones. We thus proposed that directional instruments of the type used in the graphic arts industry can also be used in the control process of textiles, especially if used with a large aperture.

In this publication, the candidate contributed to the conception of the research idea and design of the experiment together with all co-authors. Moreover, he supervised the processes of data acquisition, data analysis and writing.

3.2.2 Article G: Color appearance processing using iccMAX

Max Derhak, Phil Green, and Michele Conni. Color appearance processing using iccmax. *Electronic Imaging*, 2018(16):323–1, 2018

In this paper, we investigated some aspects of sensor adjustment transforms (SATs) using iccMAX, a new specification of the ICC colour management architecture that allows the user to have more flexibility in the creation of ICC profiles. An SAT is a mathematical procedure used to predict the appearance of a known colour coordinate after a change in observing conditions. In this paper, we have considered two types of SATs: the chromatic adaptation transforms (CATs), which are used to predict the corresponding colour for a given tristimulus value when the chromaticity of the adapting illumination changes, and the material adjustment transforms (MATs), which, unlike the CATs, assume that the material, rather than the colour, remains the same (i.e., that the reflectance of the object is constant). Therefore, whereas CATs can only address changes in illuminant, MATs are also able to pre-

dict the effect of a change of observer. In this article, we intended to show that it is possible to encode both these approaches in an iccMAX profile, and we used this capability to compare the precision of three CATs (CAT02, CAT16 and the linearized Bradford transform) and a MAT (the Wpt transform) by computing with them various changes of illuminant.

In order to achieve these goals, we encoded each SAT in an ICC profile using the iccMAX Reference Implementation. For the comparison, we employed a set of spectral reflectances obtained from the Munsell Book of Colour and from ISO 17231-1. We converted these reflectances to XYZ tristimulus values through various PCC profiles obtained by modifying those included in the Reference Implementation. The conversion from source XYZ to destination XYZ was then encoded for every SAT in the DToB `multiProcessingElements` tag of as many profiles. For each transformation, we eventually calculated the mean CIELAB ΔE^*_{ab} differences between XYZ values predicted from the reference illuminant chromatically adapted to the test illuminant, and the exact XYZ values for the test illuminant computed directly from the spectral reflectances.

The results show that the Munsell data are best predicted by the Wpt MAT, though CAT02 and CAT16 achieved a comparable accuracy. The fact that the MAT performs well was foreseeable, given that it was optimized for the Munsell reflectances. On the other hand, no single SAT performed best at predicting the change in colorimetry of the ISO 17321-1 reflectances, although the linearized Bradford transform achieved a reasonably good result.

The candidate contributed to this work by designing and implementing the experiment performed (i.e., he developed the iccMAX profiles that carried out the SATs and analysed the resulting data).

3.2.3 Article H: A versatile multi-camera system for 3d acquisition and modeling

> Oswald Lanz, Fabian Sottsas, Michele Conni, Marco Boschetti, Erica Nocerino, Fabio Menna, and Fabio Remondino. A versatile multi-camera system for 3d acquisition and modeling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:785–790, 2020

This article describes an attempt at building a market-ready multi-camera image-based 3D scanner and at developing its associated 3D digitization algorithm, real-

ized with photogrammetric reconstruction techniques. The system has been realized as a joint work between industrial and academic partners, in order to achieve the best performance possible. This setup consists of 31 synchronized high-end professional cameras fixed on a rigid structure, with an adjustable moving basement and an adaptable FOV. It can thus deal with small and medium objects, with a maximum volume of approximately $500 \times 500 \times 500\, mm^3$. The paper thoroughly describes the measurement setup, discusses its geometric and radiometric calibration and outlines the 3D reconstruction approach adopted.

The candidate collaborated by designing and performing the radiometric calibration of the device. This was achieved by implementing a target-based characterization procedure for each camera of the system. Such a process permits the acquisition of a mathematical model that transforms the raw RGB data of each camera into a device-independent colour space (CIELAB, in this case). This was performed by capturing an image of a colour target (an IT8.7/4 colour chart with 1617 patches printed on a ProofMaster Matt 140g paper) with known CIELAB coordinates with every camera. From these, we interpolated a multi-dimensional look-up table defined on a $33 \times 33 \times 33$ grid in the raw RGB space of the sensor.

3.2.4 Oral contributions

Throughout this Ph.D., the candidate participated not only in academic events, but also industrial ones. In this context, he maintained relations with the ICC [123], the Fogra Forschungsinstitut für Medientechnologien [85] and the Covision Lab [49]. On some of these occasions, he presented various parts of the work in public, in line with the application focus of an industrial Ph.D. project. Notably, the candidate held a talk at the Fogra institute in 2019 and three talks at the 2018, 2019 and 2020 ICC Developers Conferences (or DevCons). These are high-level conferences for industrial experts and academic researchers working in colour management.

For the purpose of this thesis, it is worth mentioning some details of the presentation given at the ICC DevCon 2020. The theme of ICC DevCon 2020 was "the Future of Colour Management", and it focused on implementation of ICC's new colour management architecture iccMAX. The DevCon organisers assembled a program of experts with experience of particular aspects of iccMAX and its applications. In the presentation, the candidate showed how one could use iccMAX to produce an ICC profile able to correct the spectrophotometric measurements of a sample with known topographical properties. This was achieved by implementing the simplified model of fabrics reflectance proposed in Ref. [169], which infers the

colour of the textile fiber based on the characteristics of the measurement device, the BRDF of the surface material and the height map of the sample. During the presentation, the candidate explained how to apply this model to an ideal case and demonstrated how to perform this correction with an iccMAX profile. Therefore, this was a practical application of the work performed in **Article G**. Please note, however, that this presentation was not peer-reviewed, and it should be expanded with novel concepts in order to proceed to a proper publication.

Chapter 4

Discussion

In Chapter 3, we have summarized the articles included in this dissertation, specifying motivations, methodologies and results. The current chapter discusses the outcomes of each article in the context of the research objectives listed in Sec. 1.2 and of the overall contributions to the field of texture appearance. It is divided in two parts: part one (Sec. 4.1) addresses the three research goals described in Sec. 1.2 and the relative questions, part two (Sec. 4.2) highlights the contributions of the research made.

4.1 Research goals

In this section we discuss how we approached the three research objectives introduced in Sec. 1.2 (Subsec. 4.1.1, 4.1.2 and 4.1.3) and the conclusions we can draw for each of these objectives, based on the results of the various articles. Additional remarks deducible from the publications have been gathered in Subsec. 4.1.4.

4.1.1 Study the physical requirements of a texture measurement device

In Sec. 2.3 we have summarized the possible ways in which one can physically address the measurement of texture appearance. This problem can be approached by measuring either the reflectance, whose connotation varies depending on the dimensions of the GRF considered (e.g. spatial, spectral, angular, etc.), or the 3D structure of a surface. Given the wide literature in the context of texture analysis with computer vision and the need to replicate the HVS structure, with the first goal described in Sec. 1.2 we addressed the problem of measuring the reflectance of a sample in the spectral and spatial dimensions. This issue is particularly im-

portant because, as we have shown in Article F, the complexity of a surface affects the accuracy of its colour measurement. The spectral measurement of a texture can be carried out using multi- or hyper-spectral measurement devices. The essential difference between the two methods is that hyper-spectral imaging employs a number of different wavelength channels substantially higher than multi-spectral imaging [195]. Having a large number of spectral channels allows to precisely evaluate the spectral characteristics of a material, without loss of information [178, 38]. However, extending texture analysis approaches to multiple channel images poses various problems. First of all, the computational complexity of the feature extraction algorithms increases with the amount of channels available [88]. Moreover, considering features with a high number of dimensions leads to the so-called *curse of dimensionality*, i.e. an increase in the sparsity of information, which has been observed in many pattern recognition applications [98, 225]. In the context of classification, this is also known as the *Hughes phenomenon*, and it corresponds to the reduction of mean classification accuracy when the proportion between training instances and number of spectral features is too small [113, 239, 230]. Usually, this problem is addressed either by reducing the information quantity [117, 216], i.e. by projecting the data on a subspace with a smaller number of dimensions, or by trimming some channels [224], i.e. selecting the dimensions that provide the most significant features. All these solutions, however, focus on minimizing the computational complexity of the task rather than on optimizing the physical measurement system. In **Articles A** and **B**, we tried to find the optimal hardware setup for a texture analysis system, as required by the first goal of this thesis.

In this context, Article A [45] studied how precise the spectral resolution of a snapshot device should be to correctly characterize a texture. Given its relevance and its straightforward interpretation from the point of view of perception [228], we adopted a classification framework to quantify the usefulness of spectral information. Many papers demonstrated that the accuracy of texture classification can be improved by considering not only spatial, but also colour information [194, 141, 95] and by optimizing the colour process used [33]. Based on these publications, **Article A** established the dependence of the average texture classification accuracy on the number of channels of the instrument. The results have been derived for four popular hand-crafted texture features (GLCM, LBP, Gabor and wavelet) and two widely used classifiers (3-NN and random forest). For every feature and classifier the accuracy appears to increase monotonically as $a - b \cdot e^{-c(n_{chan}-1)}$, where n_{chan} represents the number of spectral channels used, a is the maximum accuracy obtainable with the feature, $a - b$ is the accuracy obtained with just one channel and c is the exponential coefficient, which indicates how fast the function

grows. **Article A** demonstrates that for the four typologies of features used the maximum accuracy is always reached with less than ten channels. Such a number falls within the scope of multi-spectral imaging, in which the selection of optimal filters is a recurring topic [187, 106, 6, 144]. Its practical implementation could be achieved with either a multi-spectral filter wheel or a multi-spectral filter array (MSFA) [101]. Both methods have their pros and cons: a filter wheel camera has to acquire a number of images equal to the number of filters, and it is therefore slower than a MSFA-based design, which provides snapshot images. However, the latter method requires demosaicing, which negatively affects the spatial resolution of the system [183]. A possible way to avoid this limitation is to use texture features taken directly from raw MSFA images [167]. Ref. [38] points out that using multi-spectral devices instead of hyper-spectral ones has the disadvantage of not being suited for the development of a stable metric, since the measured quantities are detached from the physical characteristics of the sample. However, most reflectance functions are smooth, so hyper-spectral systems are generally redundant for perception-related purposes [6]. Another issue with the approach used in **Article A** is that the number of channels obtained with optimization analyses depends on the dataset used [38]. We tried to minimize this effect by acquiring a large image database mainly composed of fabric samples (more than 80%). The results of Article A should therefore be valid for fabrics. Another practical limitation of the paper's conclusions is that for the analysis we used ideal Gaussian filters, and their implementation with real-world materials could be difficult to achieve [152, 162]. However, various frameworks to address the problem exist, such as filter selection algorithms [217]. Moreover, the fact that we have simulated the filters instead of applying selection or transformation algorithms has allowed us to work with wide spectral transmittances. These are closer to reality, as colour filters are used for imaging applications instead of dichroic filters (which have a sharper spectral dependence) because of their transmittance is angle-independent [84].

We already argued that the precise colour reproduction of digital RGB cameras can be achieved only when their scanning filters are nonsingular transformations of the CIE CMFs [242] (the Luther-Ives condition). Nonetheless, the colorimetric calibration of a digital still camera (DSC) can be used to find the mathematical transformation from its raw RGB coordinates to a standard colour space, e.g. CIE XYZ [128]. In Article B [47] we tried to understand if the texture classification accuracies obtained with the simulated multi-spectral devices in **Article A** can be matched by simply calibrating a trichromatic camera. First of all, we compared the effect of the use of spectral sensitivities of a real-world commercial camera, a Raspberry Pi V2 [204], with those adopted in **Article A**, i.e. a set of ideal

Gaussian functions defined in the visible spectrum. We thus measured the camera CMFs and repeated the classification experiment of **Article A**. The average accuracy obtained with the Raspberry Pi sensitivities shows a feeble improvement with respect to the simulated ones. We then reproduced a complete camera acquisition pipeline and evaluated the effect of various procedural steps (calibration, rendering and sharpening). Among these, the only procedure that significantly increased the classification efficiency was rendering, which on average led to an increase of 3% for the 3-NN accuracy and of 4.5% for the RF one. In order to understand this effect, we derived the contrast of the images before and after the rendering and found a significant raise. In fact, Fig. 7 of **Article B** shows that the rendering process was based on an increment of saturation of the image colours. We can hence attribute the positive effect of rendering to the increase in the average image quality that it produces [29]. For all texture features and both classifiers, the accuracy of rendered images classification is equal to or higher than the asymptotic values derived in **Article A**. We can therefore conclude that having a high image quality is more valuable than using a radiometrically calibrated imaging system when it comes to texture classification. Note that both the calibration and the rendering steps have been implemented in **Article B** as linear transformations in a trichromatic space. However, since the texture features used for the classification are not necessarily linearly related to the pixel levels of the image from which they have been extracted, their effect on the space of the classification data is non-linear.

In other articles included in this thesis, we addressed the measurement of topographical texture. Article H [150], in particular, focuses on the general problem of the acquisition of surface topography, and discusses how to accurately reproduce the colour of a 3D object. This has been achieved by chromatically calibrating each camera of the system separately with a reference checker. We borrowed this approach from cultural heritage literature [89], since it is the main field, as far as we know, in which the issue has been studied. Additionally, we addressed the subject of texture topography in Article E [46], in which we acquired a set of 22 different fabric samples with an xTex A4 scanner [246]. This device is able to capture the appearance attributes needed for the Physically Based Rendering (PBR) [199] of a surface texture. The measurement output consists in as a set of maps (base colour, roughness, metalness, transparency, normal, displacement), compatible with the Unified 3D Material standard [236], with which one can accurately render a surface. The topographical data extracted by this setup can moreover be handily attached to an image file through an iccMAX profile, which we presented and discussed in Article G [61]. In **Article E** we evaluated the Spearman rank-correlation between the frequency at which a semantic descriptor is used

to define a texture and a set of texture features extracted from the representations of the fabric samples provided by the xTex. The results of this experiment show that for many descriptors (e.g., *shiny*, *soft* and *natural*) the highest correlation can be obtained from the normal map of the samples, suggesting that the 3D structure is able to provide useful information on the texture appearance of textiles.

4.1.2 Clarify limits and capabilities of the mathematical approaches used to study texture

The second objective proposed in Sec. 1.2 concerns the methods used for texture analysis from a more technical point of view. We thus put aside the examination of texture measuring devices and concentrated on the purely mathematical side of the texture of grey-scale images. In this regard, a careful study of the literature pointed out the lack of a standard metric for texture homogeneity (sometimes referred to as uniformity). An image is called homogeneous if its content is visually and physically inseparable. This property is fundamental for understanding texture perception, as it is connected to the studies on texture discrimination (see Sec. 2.1). Indeed, the concept of homogeneity is embedded in the field of segmentation, whose purpose is to divide an image into regions "uniform and homogeneous with respect to some characteristic such as gray tone or texture" [105]. The HVS also uses homogeneity to infer information on a surface's 3D structure [206], which is the basic idea behind shape from texture procedures [171, 86]. The definition of homogeneity has been widely studied in the context of landscape ecology [145, 158, 80], without ever having achieved a satisfying solution. In the case of texture, a widely used concept corresponding to that of homogeneity is *visual stationarity*, which has been introduced in Ref. [198]. The authors of this publication state that "a stationary texture image is an image which contains a single type of texture". However, a purely statistical interpretation of this feature also exists, as discussed in Subsec. 2.4.4. This concept of (weak) stationarity, which we have called *data stationarity* for clarity, is an important element of texture analysis because it is a fundamental hypothesis for the application of global feature extraction methods to an image [231]. In fact, if the statistical properties of the image varied depending on the position considered, a global feature, such as the average or the standard deviation of the intensity values, would provide only a partial description of the data content. While data stationarity is useful for computer science purposes, visual stationarity falls within the scope of visual appearance modeling. Although they are distinct, these two types of stationarity are often assumed to coincide in computer vision articles. This supposition is valid for textured images that satisfy the first Julesz conjecture [137], but cannot be applied

to every texture pattern [139]. Therefore, **Articles C** and **D** contain a thorough study of the relationship between data and visual stationarity.

The first step of our research consisted of choosing a spatial stationarity metric, which we selected among the tests reviewed in Subsec. 2.4.4. The LS2W-based method proposed by Taylor and colleagues in Ref. [231] appears to be the most general and stable of these techniques, and it has already been applied to computer vision in the context of image classification [232]. An interesting feature of this test is that it makes use of texture information at every dyadic scale available. As pointed out in Ref. [80], scale-dependence is a fundamental trait of texture homogeneity, because a complex scene can appear heterogeneous from close or far enough. Ref. [78] hints at this by including texture scales in a table of expanded texture appearance properties. The texture perception of an observer, in fact, varies with the viewing distance, in relation to their visual acuity and contrast sensitivity functions [159]. Due to this, some texture features have been expanded to take into account multiple scales at once [170, 248], while some others, such as the fractal dimension and granulometry [198], aim at describing the dependence on scale of a pattern. In Article C [43] we thus introduced a variation to the test proposed in Ref. [231] that analyses the stationarity of an image scale by scale. We then performed a scoping study to verify that this novel procedure worked as expected and also to study its behaviour at different scales. To do this, we first generated two sets of 100 artificial images: one based on an uncorrelated Gaussian white noise process, the other on a 2D correlated and scale-invariant Gaussian random field. We then applied the LS2W test and its scale-dependent variation to each image, calculating the average results for each group. The resulting vectors of p-values $\mathbf{p}$ are shown in Fig. 2 of the paper: both groups have an average p-value higher than the threshold $\alpha = 5\%$ at every scale. This confirms that the proposed approach correctly categorizes images which are stationary by construction. For the presentation of the paper at the LIM conference, we additionally implemented the same process with two other sets of 100 images generated through the non-stationary models NS2 and NS4 defined in Ref. [231]. The left half-plane of the NS2 images (see Fig. 4.1(a)) was filled with a unit variance white-noise process, while its right one with an isotropic Gaussian random field (Matérn covariance with shape parameter $\nu = 1$). The NS4 model, on the other hand, is a white-noise process whose standard deviation follows a sigmoidal trend in the horizontal direction, thus showing a smooth transition (see Fig. 4.1(b)). The resulting $\mathbf{p}$ vectors are reported in Fig. 4.1(c), and they show that in this case, the non-stationarity behaviour appears to be dictated by the lower scales, while higher ones are classified as stationary. This could be due to the fact that higher scales are less sensible to model changes

because the wavelets with which the p-value has been calculated spread over a wider area. Nonetheless, the average p-value of these two groups of images was smaller than the threshold, so they were rightfully identified as non-stationary. Finally, we applied the two stationarity tests to a collection of natural images drawn from the Amsterdam Library of Textures (ALOT) [28]. For four of these, the results are reported in Tab. 1 and shown in Fig. 4 of **Article C**. Here, one can observe that the ALOT image 155 has a vector $\mathbf{p}$ very similar to those of NS2 and NS4, and its overall p-value is hence 0.01, i.e., smaller than α. Image 185, on the other hand, has a p-value of 1, although it has the most clearly directional pattern and most of its vertical p-values are beneath the threshold value.

In practice, a visual stationarity metric could be used to model the perception of texture homogeneity. Therefore, Article D [44] starts from the results of **Article C** to analyse the chosen data stationarity test more in depth and to gather information on its link with visual stationarity. First of all, we extracted the vector $\mathbf{p}$ from an artificial checker image, which, as expected, had only unitary p-values. The addition of white noise with increasing standard deviation did not alter the output of the test. The noise was, in fact, stationary, therefore it did not influence the stationarity of the image. Inspired by the peculiar response of the ALOT image 185, we then applied various transformations to the chessboard picture. Initially, we stretched it in the horizontal and vertical directions, doubling the pixel length of the square primitive each time. The resulting p-values were again all unitary, apart from a small number of middle scales whose stationarity dropped for the penultimate stretch prior to reaching the maximum (see Fig. 2 in **Article D**). Since the applied stretches were global transformations, this occurrence suggests that the adopted stationarity test is subject to aberrations, though relatively negligible. The second distortion tested was the resizing of the texture primitive. An alternative interpretation of this transformation is a zooming in on a chessboard subregion. In this case, the general p-value is degraded by the distortion. Such an effect can be linked to the dependence of texture appearance on the scale discussed in Ref. [80]: as we zoom in, our perception gradually moves away from the concept of homogeneous texture and begins to consider the texture primitive as an element in its own right. In the case considered, however, this does not make the image non-stationary, because the primitive texture itself is stationary. In the second analysis of the article, we tested the assumption expressed in Ref. [198] that stationary images are better described by global texture features. In order to investigate this, we applied the concept of data stationarity to texture classification, expecting that the images classified with a mix of local and global feature vectors based on the LS2W $\mathbf{p}$ vector would achieve higher classification accuracies than using pure

feature vectors. This expectation was based on the idea the p-value of the test can be viewed as a binary criterion to decide whether an image is homogeneous or not, such as the binary criterion $\gamma(R_i)$ defined in Ref. [132]. According to the results presented in **Article D**, in which we performed a classification experiment on a subset of the DTD image database, such an hypothesis appears to be valid only for classification tasks that use human-defined groups. This suggests that the test is actually effective in reproducing the HVS mechanisms. In the final step of the study, we directly probed the relationship between data and visual stationarity by performing a visual experiment that made use of the stationary texture definition given in Ref. [198]. In fact, we asked 93 observers to group together the DTD images that looked similar. From this, we derived how similar each couple of texture images used in the classification were. We then rearranged this into a similarity map, comparing it to the confusion matrices of the second classification task with the Spearman rank-correlation. The results of this operation show which mixed feature behaves in a way analogous to the observers. In this case, a mix of GLCM and LBP and one of GLCM and wavelet features accomplish the highest value of 0.53. Therefore, the mixing of features based on stationarity data could also be useful to model the appearance of textures.

4.1.3 Gather data on textile texture perception

It has been observed that the appearance of a fabric heavily influences the choice of costumers when buying garments and apparel [227]. As humans frequently communicate perceived visual properties of materials, the most important cues that make-up appearance can be studied from a semantic point of view [77]. In Article E [46], therefore, we tried to fulfill the final goal of this thesis: the definition of a semantic context for textile appearance. This idea is not novel, as many works sook to estimate the space of texture appearance based on semantic experiments, e.g., [203, 18, 109, 97, 39, 133]. However, these experiments focus on images rather than real objects, an approach that has recently been questioned [93]. Moreover, these studies were all aimed at deriving a general texture model, disregarding the fact that humans use distinct metrics to judge different types of materials [119]. Ref. [179] solves this problem by concentrating on fabric texture semantics. In particular, by asking a group of non-expert observers to verbally describe a set of samples, the authors of this paper were able to derive a set of perceptual attributes used by people to interpret and evaluate the texture properties of textiles.

In **Article E**, on the other hand, we adopted the approach to total appearance analysis proposed by Hutchings [118], who made use of panels of experts to define

and analyse appearance in different contexts [120]. According to his method, a panel is supposed to meet, discuss and come up with a set of parameters that can be used to evaluate the appearance of a product. Hutchings mainly focused on food appearance, but the employment of expert opinions is also currently practiced in the textile industry [125]. In fact, since trained individuals have a strong grasp of the widest possible range of semantic descriptors in their field, they are more capable of providing objective observations less dependent on the social environment [148]. One of the essential abilities for an industrial designer, for example, is to "make his or her intentions explicit – communication is at the heart of industrial design" [66], although this is not always achieved [71]. Therefore, in **Article E** we gathered a set of semantic descriptors defined by 28 experts in texture design. We asked the observers to specify a maximum of five adjectives for each sample contained in a booklet comprised of 22 patches of textile. The participants were not provided with an explicit definition of texture itself. The experts were required to use the English language, so that the results could have been compared more easily, and to make use only of visual cues, given that texture is sometimes intended from a tactile point of view [74]. Based on the experimental data, we then composed a vocabulary for fabric appearance. In order to follow the principles of semantic differential theory [192], we only employed bipolar descriptors. These can be seen in Tab. 2 of **Article E**, ordered from most to least mentioned by the observers in the experiment. Among these, some can be related to mechanical properties of the samples (*stiff* and *stretchy*, *soft* and *hard*), some others to subjective aesthetic parameters (*elegant* and *cheap*, *comfortable* and *uncomfortable*). The fact that many observers used terms related to technical properties of the samples rather than visual elements prompted us to propose a semantic ontology for textiles in which the appearance can be included. In fact, an ontological framework permits us to formalize the relationship between manufacturing characteristics and appearance properties of a textile element. This could then be used to establish an appearance measurement pipeline for fabrics. After this, we evaluated the correlation between the frequency of the observers' semantic descriptors and the value of various computer vision texture features. The features have been extracted from two types of sample reproductions: a simple RGB image and an .u3m file, measured with an xTex A4 device [246], containing a total of six texture maps that can be used for PBR. The combinations of feature and map that provides the highest Spearman rank-correlation are reported in Tab. 3 of the article for each visual descriptor. It can be seen that, among the features calculated, the ones obtained from pre-trained CNNs maximize the correlation, suggesting that deep learning should be able to reproduce appearance perception. However, the neural

networks used derived 1000 features, whereas the man-made features were only 48. Moreover, while CNNs provide optimal results in computer vision [177], their main flaw is that their features are hard to explain [245], a characteristic necessary for the development of a model. To show the possible application of a texture appearance model, we finally used the features with the best correlation to predict which descriptors an expert would use when looking at a fabric image.

4.1.4 Additional Results

Let us now briefly discuss the minor contributions summarized in Secs. 3.2 and 3.2.4. Among these works, **Articles F** and **H** fall in the framework of appearance measurement, as the former addresses the effect of topographical texture on colour appraisal and the latter describes how to chromatically calibrate a 3D scanner. The first problem is very current and pressing, both in academia and industry, and remains unresolved. The standard approach to characterize the colour of a surface consists in measuring it with a spectrophotometer. This method, however, has been initially developed for matt materials, specifically paper. The results of Article F [59] show that employing it to measure textured surfaces such as textiles yields a higher measurement uncertainty. Various research studies have attempted to solve this problem [169, 234, 168, 160], but no accepted guideline has been proposed yet. In the presentation given at the ICC DevCon 2020 [42] (see Oral contributions and Sec. 3.2.4) we proposed using a simplified model of the surface reflectance and a measurement of its 3D structure to elicit a measurement correction coefficient for a spectrophotometer. This work shows how the calculations needed to implement such a model can be carried out with the calculator tool included in iccMAX [30]. IccMAX is also the focus of Article G, whose purpose was to show how various sensor adjustment transforms (SATs) can be performed in this framework.

4.2 Research Contributions

In this section we highlight the contributions to current knowledge achieved by this Ph.D. work.

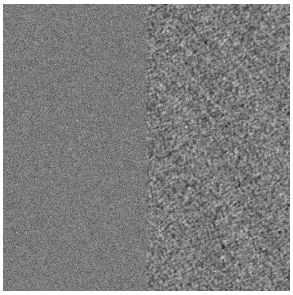
First of all, in **Article A** we defined the possibilities and limitations of spectral measurements in the analysis of textile images, showing that a multi-spectral imaging system with no more than ten spectral channels optimizes the accuracy of a texture classification pipeline. In this publication, we have expanded the study of multivariate texture analysis and classification [194, 33] to the multi- and hyperspectral cases, which is, to our knowledge, a rather unexplored field. Moreover, for

this paper we have acquired a set of 191 hyper-spectral images, which are available upon request for research purposes, 155 of which are textiles. **Article B**, on the other hand, showed that the same accuracy can be reached with a standard image acquisition pipeline. Additionally, its results suggest that a correct radiometric calibration of the system does not have a strong effect on the precision of texture classification, and that this process is therefore not needed in such a context. We have then demonstrated that, in textiles, the best performing man-made feature for textiles is LBP. Both **Article A** and **B** have been devised and carried out to meet the needs of Barbieri electronic, which was interested in acquiring information and practical data on the methods of measurement and characterization of visual texture.

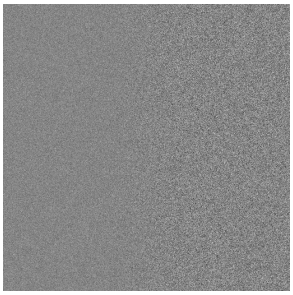
In **Articles C** and **D**, we have pointed out the dual nature of stationarity, proposing a scale-dependent method to measure it. **Article D** also showed that the division between stationary and non-stationary texture features proposed in Ref. [198] does not improve the classification accuracy in general, but can still work in some cases. This scale selection approach could be easily integrated into neural networks [193], for which the problem of bulkiness is currently widely discussed [177], to speed them up. Furthermore, **Article D** introduced a first analysis of the correlation between visual and data stationarity. All these outcomes are part of a novel approach to the study of texture discrimination.

The practical take-away from **Article E** includes a textile appearance vocabulary and a fabric ontology. Moreover, the gathered data includes not only a set of hyper-spectral images of the textile samples employed in the visual experiment, but also a joint set of maps for PBR. These results can be used as a basis to develop a standard framework for textile appearance.

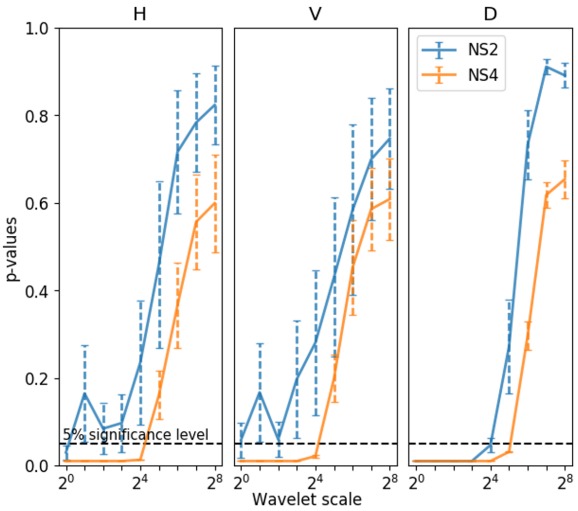
Among the minor contributions, the results of **Article F** provide useful information on the precision of colour measurement of textiles, which can be useful both for academic and industrial purposes. The analysis performed in **Article G**, on the other hand, had a dual purpose: indicate which SAT is optimal to simulate a change of illuminant and show the possible applications of iccMAX, thereby stimulating its use in both industrial and academic environments. More in general, the candidate's participation in the various ICC DevCons during the Ph.D. project (Sec. 3.2.4) was aimed at sharing ideas and innovations in the context of inter-company collaboration. Finally, **Article H** gave a complete and straightforward report on the setup of a multi-camera 3D scanner.



(a) *NS2*



(b) *NS4*



(c) *p-values*

Figure 4.1: Two images (Fig. 4.1(a) and Fig. 4.1(b)) generated with the non-stationary processes NS2 and NS4 described in Ref. [231] and a plot of the corresponding vectors $\mathbf{p}$ (Fig. 4.1(c)).

Chapter 5

Conclusions and future perspectives

In this final chapter, we provide an overall conclusion of the thesis and give a perspective for future work on the topics discussed.

5.1 Conclusions

The purpose of the Ph.D. project presented in this thesis was to advance the measurement and characterization of visual texture appearance, with special emphasis on textiles. The theory of texture appearance is still at an early stage, leaving much room for a wide range of possible research directions. In this dissertation, we have attempted to combine the optimization of current texture measurement and analysis methods with the collection of data on the mechanisms of textile perception.

First of all, we contributed to the field of texture analysis from the point of view of computer vision. In particular, we studied the dependence of texture classification on the spectral information available, and we concluded that a system optimized for texture characterization requires a number of spectral channels not bigger than ten to maximize the algorithm's accuracy. We then focused on trichromatic camera sensors, studying the effect of radiometric calibration and of common colour management processes on texture classification. The results showed that working with a scene-referred characterized camera rather than with a standard output-referred one has a relatively small effect on the accuracy of the classification. Among the steps of the digital imaging acquisition pipeline, colour rendering appears to be the

one that improves the performance of the algorithm the most, thus suggesting that rendered camera RGB images are more suited for the task than raw ones. These conclusions expand the current knowledge on spectral texture images analysis, and provide useful information for setting up a texture measurement system.

We then shifted the focus of our research on the theoretical definition of texture properties, and in particular on stationarity. In fact, we identified an ambiguity in the definition of texture stationarity, which we analysed from both the standpoint of computer vision and of appearance perception. From the computational point of view, we introduced a novel method to measure the data stationarity of an image at various scales and showed that this information can be useful to improve the accuracy of texture classification, especially when the classes have been chosen based on human visual perception. From the side of appearance, on the other hand, we gathered data on texture similarity and we calculated their correlation with the classification results. This allowed us to quantify which mix of texture features is most correlated to the human behavior. The fact that in this case, too, the mixing of some features at different spatial scales produces an improvement hints that the LS2W stationarity test can be suitable to model visual appearance. However, more in-depth studies are needed to actually derive such a model. This could be employed in the texture field to define an absolute metric of homogeneity of a visual pattern, which would greatly benefit practical operations such as image segmentation.

Finally, we contributed to the collection of data on textile appearance. We invited a group of expert observers to compile a database of semantic descriptors related to the visual appearance of textiles. These terms were then arranged into a vocabulary of dual attributes, followed by the proposal of a general ontology for fabrics. Next, we calculated the Spearman's rank-correlation between the value of a wide selection of texture features and the frequency at which the terms of the vocabulary have been used to describe each fabric samples. Among the techniques adopted, a set of pre-trained CNNs achieved the highest correlation, which is yet another confirmation of the power of neural networks in the field of computer vision. However, since their feature maps are difficult to interpret in terms understandable to humans and their outputs depend on how the network has been trained, the application of CNNs to appearance models is problematic. The fact that some of the features with high correlation have been extracted from the surface normal map showed that topographical information can contribute to define the visual texture appearance of a textile. This latest work proposes a semantic approach to the characterization of fabric appearance and provides a useful basis for further development of this

method.

5.2 Future perspectives

Sec. 4.1.1 identified the requirements that a multi-spectral system should satisfy in order to optimally classify a fabric sample, but this area can be studied further. In particular, given the rising importance of neural networks in the field of computer vision, one could study how their performance would vary with an increasing number of spectral information. Furthermore, a fundamental step in practically building a multi-spectral camera for textile classification is the choice of filters. Therefore, it would be interesting to compare the results obtained in **Article A** with ideal transmittances with real-world ones, which can be obtained from filter manufacturers [217]. On the other hand, according to the analysis performed in **Article B**, colour rendering appears to be the most important step in a trichromatic camera's image processing pipeline. A broad inquiry aimed at explaining this result would be most useful to help define an optimal texture measurement system. This could be achieved, for example, by means of a psychophysical experiment in which a group of observers chooses the most pleasant rendering setup for an imaging device and observing if and by how much their preferences affect a classification algorithm.

The difference between the concepts of data and visual stationarity, introduced in **Article C**, can be studied more in depth, too. First of all, the multiscale LS2W test proposed in the paper can be extended to the domain of multivariate images, both trichromatic and multi-spectral. A possible solution to this problem has already been developed in Ref. [232], but it is not suited for images with a high number of spectral channels because of its excessive computational load. Moreover, the adopted technique could be extended to a continuous range of scales, whereas only dyadic ones were considered in **Articles C** and **D**. Given that in **Article D** we have shown that the scale LS2W test can have some incorrect behaviour, it would also be useful to develop a wider variety of spatial stationarity metrics. In Subsec. 4.1.2 we introduced the concept of stationarity by linking it to homogeneity perception. We avoided using this term in the articles, however, because we focused only on the relationship between data and visual stationarity. It would thus be useful to assess with a visual experiment the ability of the LS2W test to quantify the homogeneity of a texture. Finally, it would be interesting to broaden the study of how the scale stationarity test behaves with regular textures beyond the chessboard case and with a wider range of distortions (e.g., the ones used in Ref. [256]).

In conclusion, the framework of fabric appearance semantics devised in **Article**

E needs to be validated [241]. This can be achieved through focused interviews with textile experts, both for the proposed vocabulary and the ontological architecture. Moreover, the model used to predict textile appearance attributes can be strengthened with rank-ordering experiments. It would also be interesting to compare the results of the psychophysical experiment performed in the paper with other types of stimuli, such as RGB images and computer graphics objects.

Bibliography

[1] Noureddine Abidi, Eric Hequet, Chris Turner, and Hamed Sari-Sarraf. Objective evaluation of durable press treatments and fabric smoothness ratings. *Textile research journal*, 75(1):19–29, 2005.

[2] Sabit Adanur. *Handbook of weaving*. CRC press, 2020.

[3] Edward H. Adelson and Alex P. Pentland. The perception of shading and reflectance. *Perception as Bayesian inference*, pages 409–423, 1996.

[4] Carlos Aliaga, Carlos Castillo, Diego Gutierrez, Miguel A Otaduy, Jorge Lopez-Moreno, and Adrian Jarabo. An appearance model for textile fibers. In *Computer Graphics Forum*, volume 36, pages 35–45. Wiley Online Library, 2017.

[5] John Aloimonos. Shape from texture. *Biological cybernetics*, 58(5):345–360, 1988.

[6] Ali Alsam, David Connah, and Jon Hardeberg. Multispectral imaging: How many sensors do we need? *Journal of Imaging Science and Technology*, 50(1):45–52, 2006.

[7] Vincent Andrearczyk and Paul F Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*, 84:63–69, 2016.

[8] Definition of *appearance* in cie's e-ilv. https://cie.co.at/eilvterm/17-22-019. accessed in May 2021.

[9] Astm e284 - 17: Standard terminology of appearance. www.astm.org.

[10] André Ricardo Backes, Dalcimar Casanova, and Odemir Martinez Bruno. Color texture analysis based on fractal descriptors. *Pattern Recognition*, 45(5):1984–1992, 2012.

[11] I. K. Baldry and J. Bland-Hawthorn. A tunable echelle imager. *Publications of the Astronomical Society of the Pacific*, 112(774):1112–1120, 2000.

[12] Martin S. Banks, Jenny C. A. Read, Robert S. Allison, and Simon J. Watt. Stereoscopy and the human visual system. *SMPTE motion imaging journal*, 121(4):24–43, 2012.

[13] Bryce E. Bayer. Color imaging array, 1976. US Patent 3,971,065.

[14] Raquel Bello-Cerezo, Francesco Bianconi, Francesco Di Maria, Paolo Napoletano, and Fabrizio Smeraldi. Comparative evaluation of hand-crafted image descriptors vs. off-the-shelf cnn-based features for colour texture classification under ideal and realistic conditions. *Applied Sciences*, 9(4):738, 2019.

[15] James R Bergen and Bela Julesz. Parallel versus serial processing in rapid pattern discrimination. *Nature*, 303(5919):696–698, 1983.

[16] Roy S Berns. *Billmeyer and Saltzman's principles of color technology*. John Wiley & Sons, Inc., 2019.

[17] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24:192–236, 1974.

[18] Nalini Bhushan, A Ravishankar Rao, and Gerald L Lohse. The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images. *Cognitive science*, 21(2):219–246, 1997.

[19] Griff Bilbro. Technology options for multi-spectral infrared cameras. Technical report, North Carolina State University at Raleigh Department of Electrical and Computer Engineering, 2001.

[20] Rémy Blanc, Jean-Pierre Da Costa, Youssef Stitou, Pierre Baylou, and Christian Germain. Assessment of texture stationarity using the asymptotic behavior of the empirical mean and variance. *IEEE transactions on image processing*, 17(9):1481–1490, 2008.

[21] Liam Blunt and Xiang Jiang. *Advanced techniques for assessment surface topography: development of a basis for 3D surface texture standards*. Elsevier, 2003.

[22] Giulia Boato, Duc-Tien Dang-Nguyen, and Francesco GB De Natale. Morphological filter detector for image forensics applications. *IEEE Access*, 8:13549–13560, 2020.

[23] Sandip Bose and Allan O Steinhardt. Invariant tests for spatial stationarity using covariance structure. *IEEE transactions on signal processing*, 44(6):1523–1533, 1996.

[24] Sheryl Brahnam, Lakhmi C Jain, Loris Nanni, Alessandra Lumini, et al. *Local binary patterns: new variants and applications*. Springer, 2014.

[25] Johannes Brauers and Til Aach. Geometric calibration of lens and filter distortions for multispectral filter-wheel cameras. *IEEE Transactions on Image Processing*, 20(2):496–505, 2010.

[26] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

[27] Theodor V. Bulygin and Gennady N. Vishnyakov. Spectrotomography: a new method of obtaining spectrograms of two-dimensional objects. In *Analytical Methods for Optical Tomography*, volume 1843, pages 315–323. International Society for Optics and Photonics, 1992.

[28] Gertjan J Burghouts and Jan-Mark Geusebroek. Material-specific adaptation of color invariant features. *Pattern Recognition Letters*, 30(3):306–313, 2009.

[29] Norman Burningham, Zygmunt Pizlo, and Jan P Allebach. Image quality metrics. *Encyclopedia of imaging science and technology*, 1:598–616, 2002.

[30] iccmax multiprocessingelement calculator programming. https://www.color.org/whitepapers/ICC_White_Paper45_Calculator_Programming-v3.pdf. accessed in August 2021.

[31] Susan Carden. *Digital textile printing*. Bloomsbury Publishing, 2015.

[32] Carlos Castillo, Jorge López-Moreno, and Carlos Aliaga. Recent advances in fabric appearance reproduction. *Computers & Graphics*, 84:103–121, 2019.

[33] Eva Cernadas, M Fernández-Delgado, Encarnación González-Rufino, and Pilar Carrión. Influence of normalization and color space to color texture classification. *Pattern Recognition*, 61:120–138, 2017.

[34] Alice C Chadwick and RW Kentridge. The perception of gloss: A review. *Vision research*, 109:221–235, 2015.

[35] Jyotismita Chaki and Nilanjan Dey. *Texture feature extraction techniques for image recognition*. Springer, 2020.

[36] MICHAEL J Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proceedings-Vision, Image and Signal Processing*, 142(4):199–206, 1995.

[37] Asim Kumar Roy Choudhury. *Principles of colour and appearance measurement: Object appearance, colour perception and instrumental measurement*. Elsevier, 2014.

[38] Rui Jian Chu, Noël Richard, Hermine Chatoux, Christine Fernandez-Maloigne, and Jon Yngve Hardeberg. Hyperspectral texture metrology based on joint probability of spectral and spatial distribution. *IEEE Transactions on Image Processing*, 30:4341–4356, 2021.

[39] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3606–3613, 2014.

[40] Mircea Cimpoi, Subhransu Maji, and Andrea Vedaldi. Deep filter banks for texture recognition and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3828–3836, 2015.

[41] M Clark, A Bovik, and W Geisler. Texture segmentation using a class of narrowband filters. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87*, volume 12, pages 571–574. IEEE, 1987.

[42] Michele Conni. iccMAX in Barbieri. https://www.color.org/DevCon/devcon2020, 2020. Presented at ICC DevCon 2020: The Future of Color Management.

[43] Michele Conni and Hilda Deborah. Texture stationarity evaluation with local wavelet spectrum. In *London Imaging Meeting*, volume 2020, pages 24–27. Society for Imaging Science and Technology, 2020.

[44] Michele Conni, Hilda Deborah, Peter Nussbaum, and Phil Green. Visual and data stationarity of texture images. *Journal of Electronic Imaging*, 30(4):043001, 2021.

[45] Michele Conni, Helene Midtfjord, Peter Nussbaum, and Phil Green. Dependence of texture classification accuracy on spectral information. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018.

[46] Michele Conni, Peter Nussbaum, and Phil Green. Textile texture descriptors. *Manuscript under review in a journal*.

[47] Michele Conni, Peter Nussbaum, and Phil Green. The effect of camera calibration on multichannel texture classification. *Journal of Imaging Science and Technology*, 65(1):10503–1, 2021.

[48] Georges Courtes. Methodes d'observation et étude de l'hydrogéne interstellaire en émission. In *Annales d'Astrophysique*, volume 23, page 115, 1960.

[49] Covision lab website. https://www.covisionlab.com/en. accessed in August 2021.

[50] George R. Cross and Anil K. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(1):25–39, 1983.

[51] Tamás Czimmermann, Gastone Ciuti, Mario Milazzo, Marcello Chiurazzi, Stefano Roccella, Calogero Maria Oddo, and Paolo Dario. Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, 20(5):1459, 2020.

[52] Xiyang Dai, Joe Yue-Hei Ng, and Larry S Davis. Fason: First and second order information fusion network for texture recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7352–7360, 2017.

[53] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

[54] Ingrid Daubechies. *Ten lectures on wavelets*. SIAM, 1992.

[55] John G Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *JOSA A*, 2(7):1160–1169, 1985.

[56] ER Davies. Introduction to texture analysis. In *Handbook of texture analysis*, pages 1–31. World Scientific, 2008.

[57] Peter de Groot. Coherence scanning interferometry. In *Optical measurement of surface topography*, pages 187–208. Springer, 2011.

[58] Peter de Groot. Phase shifting interferometry. In *Optical measurement of surface topography*, pages 167–186. Springer, 2011.

[59] Nadile Nunes de Lima, Michele Conni, Phil Green, and Markus Barbieri. Measurement uncertainty for printed textiles. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018.

[60] Hilda Deborah and Noël Richard. Morphological texture description for hyperspectral images: Pattern spectrum. 03 2021.

[61] Max Derhak, Phil Green, and Michele Conni. Color appearance processing using iccmax. *Electronic Imaging*, 2018(16):323–1, 2018.

[62] Max Derhak, Phil Green, and Tom Lianza. Introducing iccmax: new frontiers in color management. In *Color Imaging XX: Displaying, Processing, Hardcopy, and Applications*, volume 9395, page 93950L. International Society for Optics and Photonics, 2015.

[63] Maxim W Derhak and Roy S Berns. Introducing wpt (waypoint): A color equivalency representation for defining a material adjustment transform. *Color Research & Application*, 40(6):535–549, 2015.

[64] Simon-Frédéric Désage, Gilles Pitard, Maurice Pillet, Hugues Favrelière, Jean-Luc Maire, Fabrice Frelin, Serge Samper, and Gaëtan Le Goïc. Extended visual appearance texture features. In *Measuring, Modeling, and Reproducing Material Appearance 2015*, volume 9398, page 93980K. International Society for Optics and Photonics, 2015.

[65] Michael R. Descour. Throughput advantage in imaging fourier-transform spectrometers. In *Imaging Spectrometry II*, volume 2819, pages 285–291. International Society for Optics and Photonics, 1996.

[66] Peter Dormer. *Design since 1945*. Thames and Hudson London, 1993.

[67] Julie Dorsey, Holly Rushmeier, and François Sillion. *Digital modeling of material appearance*. Elsevier, 2010.

[68] Alexandru Drimbarean and Paul F Whelan. Experiments in colour texture analysis. *Pattern recognition letters*, 22(10):1161–1167, 2001.

[69] Describable textures dataset (dtd) website. http://www.robots.ox.ac.uk/~vgg/data/dtd/. accessed in September 2020.

[70] Dennis Dunn and William E Higgins. Optimal gabor filters for texture segmentation. *IEEE Transactions on image processing*, 4(7):947–964, 1995.

[71] Claudia M Eckert, Nigel Cross, and Jeffrey H Johnson. Intelligent support for communication in design teams: garment shape specifications in the knitwear industry. *Design studies*, 21(1):99–112, 2000.

[72] Idris A. Eckley, Guy P. Nason, and Robert L. Treloar. Locally stationary wavelet fields with application to the modelling and analysis of image texture. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(4):595–616, 2010.

[73] Ray F. Egerton. *Physical principles of electron microscopy*, volume 56. Springer, 2005.

[74] Galal Elkharraz, Stefan Thumfart, Diyar Akay, Christian Eitzinger, and Brian Henson. Making tactile textures with predefined affective properties. *IEEE Transactions on Affective Computing*, 5(1):57–70, 2013.

[75] A Ephraty, J Tabrikian, and H Messer. A test for spatial stationarity and applications. In *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, pages 412–415. IEEE, 1996.

[76] A Ephraty, J Tabrikian, and H Messer. Underwater source detection using a spatial stationarity test. *The Journal of the Acoustical Society of America*, 109(3):1053–1063, 2001.

[77] Lars Eriksson, Bengt-Göran Rosen, Martin Bergman, et al. Affective surface engineering-using soft and hard metrology to measure the sensation and perception in surface properties. *DS 91: Proceedings of NordDesign 2018, Linköping, Sweden, 14th-17th August 2018*, 2018.

[78] Christian Eugène. Measurement of "total visual appearance": a cie challenge of soft metrology. In *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement*, pages 61–65, 2008.

[79] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, Inc., 2013.

[80] Rusty A Feagin. Heterogeneity versus homogeneity: a conceptual and mathematical theory in terms of scale-invariant and scale-covariant distributions. *Ecological Complexity*, 2(4):339–356, 2005.

[81] Jirí Filip, Mike J Chantler, Patrick R Green, and Michal Haindl. A psychophysically validated metric for bidirectional texture data reduction. *ACM Trans. Graph.*, 27(5):138, 2008.

[82] Jiří Filip, Martina Kolafová, Michal Havlíček, Radomír Vávra, Michal Haindl, and Holly Rushmeier. Evaluating physical and rendered material appearance. *The Visual Computer*, 34(6):805–816, 2018.

[83] Jiří Filip, Pavel Vácha, and Michal Haindl. Analysis of human gaze interactions with texture and shape. In *Internatinoal Workshop on computational Intelligence for Multimedia Understanding*, pages 160–171. Springer, 2011.

[84] Description of filters used in machine vision in the edmund optics knowledge center. https://www.edmundoptics.com/knowledge-center/application-notes/imaging/filtering-in-machine-vision. accessed in August 2021.

[85] Fogra forschungsinstitut für medientechnologien website. https://fogra.org/. accessed in August 2021.

[86] David A Forsyth. Shape from texture without boundaries. In *European Conference on Computer Vision*, pages 225–239. Springer, 2002.

[87] Montserrat Fuentes. A formal test for nonstationarity of spatial stochastic processes. *Journal of Multivariate Analysis*, 96(1):30–54, 2005.

[88] Keinosuke Fukunaga and Raymond R. Hayes. Effects of sample size in classifier design. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):873–885, 1989.

[89] Marco Gaiani, Fabrizio Ivan Apollonio, Andrea Ballabeni, and Fabio Remondino. Securing color fidelity in 3d architectural heritage scenarios. *Sensors*, 17(11):2437, 2017.

[90] Nahum Gat. Imaging spectroscopy using tunable filters: a review. In *Wavelet Applications VII*, volume 4056, pages 50–65. International Society for Optics and Photonics, 2000.

[91] Davit Gigilashvili, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Marius Pedersen. Translucency perception: A review. *Journal of Vision*, 21(8):4–4, 2021.

[92] Alexander F. H. Goetz, Gregg Vane, Jerry E. Solomon, and Barrett N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1153, 1985.

[93] Michael A Gomez, Rafal M Skiba, and Jacqueline C Snow. Graspable objects grab attention more than images do. *Psychological Science*, 29(2):206–218, 2018.

[94] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice hall Upper Saddle River, NJ, 2002.

[95] Encarnación González-Rufino, Pilar Carrión, Eva Cernadas, Manuel Fernández-Delgado, and Rosario Domínguez-Petit. Exhaustive comparison of colour texture features and classification methods to discriminate cells categories in histological images of fish ovary. *Pattern Recognition*, 46:2391, 2013.

[96] Phil Green and Lindsay MacDonald. *Colour engineering: achieving device independent colour*, volume 30. John Wiley & Sons, Inc., 2011.

[97] Xiaoying Guo, Chie Muraki Asano, Akira Asano, and Takio Kurita. Visual complexity perception and texture image characteristics. In *2011 International Conference on Biometrics and Kansei Engineering*, pages 260–265. IEEE, 2011.

[98] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1):389–422, 2002.

[99] Tanzima Habib, Phil Green, and Peter Nussbaum. Brdf rendering by interpolation of optimised model parameters. In *Color and Imaging Conference*, volume 2020, pages 162–168. Society for Imaging Science and Technology, 2020.

[100] Tanzima Habib, Phil Green, and Peter Nussbaum. Spectral reproduction: drivers, use cases and workflow. *Electronic Imaging*, 2020(15):121–1, 2020.

[101] Nathan A Hagen and Michael W Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901, 2013.

[102] Michal Haindl and Jiri Filip. *Visual texture: Accurate material appearance measurement, representation and modeling*. Springer Science & Business Media, 2013.

[103] Junwei Han, Stephen J McKenna, and Ruixuan Wang. Regular texture analysis as statistical model selection. In *European Conference on Computer Vision*, pages 242–255. Springer, 2008.

[104] Robert M. Haralick, K. Shanmugam, and Its'Hak Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621, 1973.

[105] Robert M Haralick and Linda G Shapiro. Image segmentation techniques. *Computer vision, graphics, and image processing*, 29(1):100–132, 1985.

[106] Jon Y Hardeberg. Filter selection for multispectral color image acquisition. *Journal of Imaging Science and Technology*, 48(2):105–110, 2004.

[107] Andrew R. Harvey and David W. Fletcher-Holmes. High-throughput snapshot spectral imaging in two dimensions. In *Spectral Imaging: Instrumentation, Applications, and Analysis II*, volume 4959, pages 46–55. International Society for Optics and Photonics, 2003.

[108] James E. Harvey, Narak Choi, Sven Schroeder, and Angela Duparré. Total integrated scatter from surfaces with arbitrary roughness, correlation widths, and incident angles. *Optical Engineering*, 51(1):013402, 2012.

[109] Christopher Heaps and Stephen Handel. Similarity and features of natural textures. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2):299, 1999.

[110] E. Hecht. *Optics*. Pearson Education, Inc., 2012.

[111] Paul S Heckbert. Survey of texture mapping. *IEEE computer graphics and applications*, 6(11):56–67, 1986.

[112] Ian P. Howard and Brian J. Rogers. *Binocular vision and stereopsis*. Oxford University Press, USA, 1995.

[113] Gordon Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE transactions on information theory*, 14(1):55–63, 1968.

[114] Shao Hui, Wang Jianyu, and Xue Yongqi. Key technology of pushbroom hyperspectral imager (phi). *Journal of Remote Sensing*, 2(4):251–254, 1998.

[115] Chih-Cheng Hung, Enmin Song, and Yihua Lan. Image texture, texture features, and image texture classification and segmentation. In *Image Texture Analysis*, pages 3–14. Springer, 2019.

[116] Robert William Gainer Hunt. *The reproduction of colour*. John Wiley & Sons, Inc., 2005.

[117] Lian-Zhi Huo and Ping Tang. Spectral and spatial classification of hyperspectral data using svms and gabor textures. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1708–1711. IEEE, 2011.

[118] John Hutchings. The continuity of colour, design, art, and science. i. the philosophy of the total appearance concept and image measurement. *Color Research & Application*, 20(5):296–306, 1995.

[119] John Hutchings. The continuity of colour, design, art, and science. ii. application of the total appearance concept to image creation. *Color Research & Application*, 20(5):307–312, 1995.

[120] John B Hutchings. Sensory assessment of appearance—methodology. In *Food colour and appearance*, pages 104–141. Springer, 1994.

[121] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[122] Hyspex vnir-1800 main specifications. https://www.hyspex.com/media/vygmbbgw/hyspex_vnir-1800.pdf. accessed in June 2021.

[123] International colour consortium (icc) website. https://www.color.org. accessed in August 2021.

[124] Specification ICC.2:2019 (Profile version 5.0.0 - iccMAX). http://www.color.org/specification/ICC.2-2019.pdf.

[125] Textiles — Tests for colour fastness — Part A02: Grey scale for assessing change in colour. www.iso.org.

[126] Geometrical Product Specifications (GPS) — Indication of surface texture in technical product documentation. www.iso.org.

[127] Graphic technology — spectral measurement and colorimetric computation for graphic arts images. www.iso.org.

[128] Graphic technology and photography — colour characterisation of digital still cameras (dscs) — part 1: Stimuli, metrology and test procedures. www.iso.org.

[129] Geometrical product specifications (gps) — surface texture: Areal — part 2: Terms, definitions and surface texture parameters. www.iso.org.

[130] Geometrical product specifications (gps) — surface texture: Areal — part 6: Classification of methods for measuring surface texture. www.iso.org.

[131] Graphic technology - assessment and validation of the performance of spectrocolorimeters and spectrodensitometers. www.iso.org.

[132] Mihai Ivanovici, Noël Richard, and Dietrich Paulus. Color image segmentation. In *advanced color image processing and analysis*, pages 219–277. Springer, 2013.

[133] Richard HAH Jacobs, Koen V Haak, Stefan Thumfart, Remco Renken, Brian Henson, and Frans W Cornelissen. Aesthetics by numbers: links between perceived texture qualities and computed visual texture properties. *Frontiers in human neuroscience*, 10:343, 2016.

[134] Pierre J. C. Janssen. Sur l'étude spectrale des protubérances solaires [spectral study of solar prominences]. *Comptes rendus de l'Académie des Sciences*, 68:93–95, 1869.

[135] Stefan Jansson. Evaluation of methods for estimating fractal properties of intensity images. *MTech thesis, Umea University, Umea, Sweden*, pages 172–180, 2006.

[136] Rebecca K. Jones and David N. Lee. Why two eyes are better than one: the two views of binocular vision. *Journal of Experimental Psychology: Human Perception and Performance*, 7(1):30, 1981.

[137] Bela Julesz. Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2):84–92, 1962.

[138] Bela Julesz. Experiments in the visual perception of texture. *Scientific American*, 232(4):34–43, 1975.

[139] Bela Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological cybernetics*, 41(2):131–138, 1981.

[140] Christos Kampouris, Stefanos Zafeiriou, Abhijeet Ghosh, and Sotiris Malassiotis. Fine-grained material classification using micro-geometry and reflectance. In *European Conference on Computer Vision*, pages 778–792. Springer, 2016.

[141] Umasankar Kandaswamy, Stephanie A Schuckers, and Donald Adjeroh. Comparison of texture analysis schemes under nonideal conditions. *IEEE Transactions on Image Processing*, 20:2260, 2010.

[142] Tae Jin Kang, Dae Hwan Cho, and Sung Min Kim. New objective evaluation of fabric smoothness appearance. *Textile Research Journal*, 71(5):446–453, 2001.

[143] James M Keller, Susan Chen, and Richard M Crownover. Texture description and segmentation through fractal geometry. *Computer Vision, Graphics, and image processing*, 45(2):150–166, 1989.

[144] Haris Ahmad Khan, Sofiane Mihoubi, Benjamin Mathon, Jean-Baptiste Thomas, and Jon Yngve Hardeberg. Hytexila: High resolution visible and near infrared hyperspectral texture images. *Sensors*, 18(7):2045, 2018.

[145] Jurek Kolasa and C David Rollo. Introduction: the heterogeneity of heterogeneity: a glossary. In *Ecological heterogeneity*, pages 1–23. Springer, 1991.

[146] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[147] Michael W. Kudenov, Matthew E. L. Jungwirth, Eustace L. Dereniak, and Grant R. Gerhart. White light sagnac interferometer for snapshot linear polarimetric imaging. *Optics express*, 17(25):22520, 2009.

[148] M. Kuper. The Colour of Words - Blue, Green, Black and a Dappled Grey Horse. https://stories.rosaproductions.co.uk/---the-colour-of-words---/. accessed in June 2021.

[149] Michael S Landy. Texture analysis and perception. *The new visual neurosciences*, pages 639–652, 2013.

[150] Oswald Lanz, Fabian Sottsas, Michele Conni, Marco Boschetti, Erica Nocerino, Fabio Menna, and Fabio Remondino. A versatile multi-camera system for 3d acquisition and modeling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:785–790, 2020.

[151] Pierre-Jean Lapray, Jean-Baptiste Thomas, Pierre Gouton, and Yassine Ruichek. Energy balance in spectral filter array camera design. *Journal of the European Optical Society-Rapid Publications*, 13(1), 2017.

[152] Pierre-Jean Lapray, Xingbo Wang, Jean-Baptiste Thomas, and Pierre Gouton. Multispectral filter arrays: Recent advances and practical implementation. *Sensors*, 14(11):21626–21659, 2014.

[153] K. C. Lawrence, B. Park, W. R. Windham, and C. Mao. Calibration of a pushbroom hyperspectral imaging system for agricultural inspection. *Transactions of the ASAE*, 46(2):513, 2003.

[154] Richard Leach. *Optical measurement of surface topography*, volume 8. Springer, 2011.

[155] Richard Leach. *Fundamental principles of engineering nanometrology*. Elsevier, 2014.

[156] Audrey Ledoux, Olivier Losson, and Ludovic Macaire. Color local binary patterns: compact descriptors for texture classification. *Journal of Electronic Imaging*, 25(6):061404, 2016.

[157] Frédéric B Leloup, Gael Obein, Michael R Pointer, and Peter Hanselaer. Toward the soft metrology of surface gloss: A review. *Color Research & Application*, 39(6):559–570, 2014.

[158] H Li and JF Reynolds. On definition and quantification of heterogeneity. *Oikos*, pages 280–284, 1995.

[159] Liang Li, Akira Asano, Chie Muraki Asano, and Katsunori Okajima. Statistical quantification of the effects of viewing distance on texture perception. *JOSA A*, 30(7):1394–1403, 2013.

[160] Peng-Fei Li, Yu-Wei Ning, and Jun-Feng Jing. Research on the detection of fabric color difference based on t-s fuzzy neural network. *Color Research & Application*, 42(5):609–618, 2017.

[161] Stan Z Li. *Markov random field modeling in image analysis*. Springer Science & Business Media, 2009.

[162] Yuqi Li, Aditi Majumder, Hao Zhang, and M Gopi. Optimized multi-spectral filter array based imaging of natural scenes. *Sensors*, 18(4):1172, 2018.

[163] Tsung-Yu Lin, Aruni RoyChowdhury, and Subhransu Maji. Bilinear cnn models for fine-grained visual recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 1449–1457, 2015.

[164] Li Liu, Jie Chen, Paul Fieguth, Guoying Zhao, Rama Chellappa, and Matti Pietikäinen. From bow to cnn: Two decades of texture representation for texture classification. *International Journal of Computer Vision*, 127(1):74–109, 2019.

[165] Shengyang Liu, Lei Dong, Xiaozhong Liao, Ying Hao, Xiaodong Cao, and Xiaoxiao Wang. A dilation and erosion-based clustering approach for fault diagnosis of photovoltaic arrays. *IEEE Sensors Journal*, 19(11):4123–4137, 2019.

[166] Yanxi Liu, Robert T Collins, and Yanghai Tsin. A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE transactions on pattern analysis and machine intelligence*, 26(3):354–371, 2004.

[167] Olivier Losson, Alice Porebski, Nicolas Vandenbroucke, and Ludovic Macaire. Color texture analysis using cfa chromatic co-occurrence matrices. *Computer Vision and Image Understanding*, 117(7):747–763, 2013.

[168] Lin Luo, Hui-Liang Shen, Si-Jie Shao, and John Xin. Empirical model for matching spectrophotometric reflectance of yarn windings and multispectral imaging reflectance of single strands of yarns. *JOSA A*, 32(8):1459–1467, 2015.

[169] Lin Luo, Ka Man Tsang, Hui-Liang Shen, Si-Jie Shao, and John H Xin. An investigation of how the texture surface of a fabric influences its instrumental color. *Color Research & Application*, 40(5):472–482, 2015.

[170] Topi Mäenpää and Matti Pietikäinen. Multi-scale binary patterns for texture analysis. In *Scandinavian conference on image analysis*, pages 885–892. Springer, 2003.

[171] Jitendra Malik and Ruth Rosenholtz. Computing local surface orientation and shape from texture for curved surfaces. *International journal of computer vision*, 23(2):149–168, 1997.

[172] Bangalore S Manjunath and Wei-Ying Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, 1996.

[173] Jianchang Mao and Anil K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern recognition*, 25(2):173–188, 1992.

[174] Armando Martinez, Noël Richard, and Christine Fernandez-Maloigne. Colour contrast occurrence matrix: a vector and perceptual texture feature. In *Color imaging Conference*, number 23. IS&T, 2015.

[175] Masakazu Matsugu, Katsuhiko Mori, Yusuke Mitari, and Yuji Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559, 2003.

[176] Robert K. McConnell. Method of and apparatus for pattern recognition, 1986.

[177] Gaurav Menghani. Efficient deep learning: A survey on making deep learning models smaller, faster, and better. *arXiv preprint arXiv:2106.08962*, 2021.

[178] Arash Mirhashemi. Introducing spectral moment features in analyzing the spectex hyperspectral texture database. *Machine Vision and Applications*, 29(3):415–432, 2018.

[179] Fereshteh Mirjalili and Jon Yngve Hardeberg. Appearance perception of textiles: a tactile and visual texture study. In *Color and Imaging Conference*, volume 2019, pages 43–48. Society for Imaging Science and Technology, 2019.

[180] Majid Mirmehdi. *Handbook of texture analysis*. Imperial College Press, 2008.

[181] Fardin Mirzapour and Hassan Ghassemian. Using glcm and gabor filters for classification of pan images. In *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, pages 1–6. IEEE, 2013.

[182] Takeshi Mita, Toshimitsu Kaneko, Bjorn Stenger, and Osamu Hori. Discriminative feature co-occurrence selection for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1257–1269, 2008.

[183] Yusukex Monno, Sunao Kikuchi, Masayuki Tanaka, and Masatoshi Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059, 2015.

[184] Douglas B. Murphy. *Fundamentals of light microscopy and electronic imaging*. John Wiley & Sons, Inc., 2002.

[185] Guy Nason. *Wavelet methods in statistics with R*. Springer Science & Business Media, 2008.

[186] Yasir Nawab. *Textile Engineering*. Walter de Gruyter GmbH, Germany, 2016.

[187] Gianluca Novati, Paolo Pellegri, and Raimondo Schettini. Selection of filters for multispectral acquisition using the filter vectors analysis method. In *Color Imaging IX: Processing, Hardcopy, and Applications*, volume 5293, pages 20–26. International Society for Optics and Photonics, 2003.

[188] Timo Ojala, Topi Mäenpää, Jaakko Viertola, Juha Kyllönen, and Matti Pietikäinen. Empirical evaluation of mpeg-7 texture descriptors with a large-scale experiment. In *Proc. 2nd International Workshop on Texture Analysis and Synthesis*, pages 99–102. in: Proc. 2nd International Workshop on Texture Analysis and Synthesis, Copenhagen, Denmark, 99-102, 2002.

[189] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[190] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Optics letters*, 16(16):1277, 1991.

[191] Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[192] Charles Egerton Osgood, George J Suci, and Percy H Tannenbaum. *The measurement of meaning*. Number 47. University of Illinois press, 1957.

[193] Niall O'Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep learning vs. traditional computer vision. In *Science and Information Conference*, pages 128–144. Springer, 2019.

[194] Christoph Palm. Color texture classification by integrative co-occurrence matrices. *Pattern recognition*, 37(5):965–976, 2004.

[195] Rüdiger Paschotta. Definition of *hyperspectral imaging* in the rp photonics encyclopedia. https://www.rp-photonics.com/hyperspectral_imaging.html. accessed in August 2021.

[196] Frank L. Pedrotti, Leno M. Pedrotti, and Leno S. Pedrotti. *Introduction to optics*. Cambridge University Press, 2017.

[197] Maria Petrou and Pedro Garcia Sevilla. *Image processing: dealing with texture*. Wiley, 2006.

[198] Maria M. P. Petrou and Sei-ichiro Kamata. *Image processing: dealing with texture*. John Wiley & Sons, Inc., 2021.

[199] Matt Pharr, Wenzel Jakob, and Greg Humphreys. *Physically based rendering: From theory to implementation*. Morgan Kaufmann, 2016.

[200] Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. *Computer vision using local binary patterns*, volume 40. Springer Science & Business Media, 2011.

[201] Michael R. Pointer. Measuring visual appearance - a framework for the future. 2003.

[202] William K. Pratt. *Introduction to digital image processing*. John Wiley & Sons, Inc., 2001.

[203] A Ravishankar Rao and Gerald L Lohse. Towards a texture naming system: Identifying relevant dimensions of texture. *Vision Research*, 36(11):1649–1669, 1996.

[204] Raspberry pi v2 camera documentation. https://www.raspberrypi.org/documentation/raspbian/applications/camera.md. accessed in June 2021.

[205] Ruth Rosenholtz. Texture perception. *Oxford Handbook of Perceptual Organization*, 167:186, 2014.

[206] Ruth Rosenholtz and Jitendra Malik. Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision research*, 37(16):2283–2293, 1997.

[207] Stephen J Russell. Handbook of nonwovens. 2006.

[208] Sanjoy K. Saha, Amit K. Das, and Bhabatosh Chanda. Cbir using perception based texture and colour measures. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 985–988. IEEE, 2004.

[209] Samuele Salti, Federico Tombari, and Luigi Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014.

[210] Faten Sandid and Ali Douik. Texture descriptor based on local combination adaptive ternary pattern. *IET Image Processing*, 9(8):634–642, 2015.

[211] Definition of *Semantic Web* on the w3c's website. https://www.w3.org/standards/semanticweb/. accessed June 2021.

[212] Gaurav Sharma. Color fundamentals for digital imaging. In *Digital color imaging handbook*, pages 1–114. CRC press, 2017.

[213] Gaurav Sharma and Raja Bala. *Digital color imaging handbook*. CRC press, 2017.

[214] Monika Sharma and Hiranmay Ghosh. Histogram of gradient magnitudes: a rotation invariant texture-descriptor. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4614–4618. IEEE, 2015.

[215] James E. Sheedy, Ian L. Bailey, Markus Buri, and Eric Bass. Binocular vs. monocular task performance. *American journal of optometry and physiological optics*, 63(10):839–846, 1986.

[216] Miaohong Shi and Glenn Healey. Hyperspectral texture recognition using a multiscale opponent representation. *IEEE transactions on geoscience and remote sensing*, 41(5):1090–1095, 2003.

[217] Raju Shrestha. *Multispectral imaging: Fast acquisition, capability extension, and quality evaluation*. PhD thesis, PhD thesis, University of Oslo, 2014.

[218] Thomas Smith and John Guild. The cie colorimetric standards and their use. *Transactions of the optical society*, 33(3):73, 1931.

[219] Pierre Soille. Morphological texture analysis: an introduction. In *Morphology of condensed matter*, pages 215–237. Springer, 2002.

[220] Pierre Soille. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[221] David J Spencer. *Knitting technology: a comprehensive handbook and practical guide*, volume 16. CRC press, 2001.

[222] Andrew Stockman, Donald IA MacLeod, and Nancy E Johnson. Spectral sensitivities of the human cones. *JOSA A*, 10(12):2491–2521, 1993.

[223] Definition of *sub-surface texture* in cie's e-ilv. https://cie.co.at/eilvterm/17-24-139. accessed in May 2021.

[224] Weiwei Sun and Qian Du. Hyperspectral band selection: A review. *IEEE Geoscience and Remote Sensing Magazine*, 7(2):118–139, 2019.

[225] Zehang Sun, George Bebis, and Ronald Miller. Object detection using feature subset selection. *Pattern recognition*, 37(11):2165–2176, 2004.

[226] Definition of *surface texture* in cie's e-ilv. https://cie.co.at/eilvterm/17-24-140. accessed in May 2021.

[227] Mary E Swinker and Jean D Hines. Understanding consumers' perception of clothing quality: A multidimensional approach. *International journal of consumer studies*, 30(2):218–223, 2006.

[228] Hideyuki Tamura, Shunji Mori, and Takashi Yamawaki. Textural features corresponding to visual perception. *IEEE Transactions on Systems, man, and cybernetics*, 8(6):460–473, 1978.

[229] Mark Tapia. A visual implementation of a shape grammar system. *Environment and Planning B: Planning and Design*, 26(1):59–73, 1999.

[230] Gülşen Taşkın, Hüseyin Kaya, and Lorenzo Bruzzone. Feature selection based on high dimensional model representation for hyperspectral images. *IEEE Transactions on Image Processing*, 26(6):2918–2928, 2017.

[231] Sarah L. Taylor, Idris A. Eckley, and Matthew A. Nunes. A test of stationarity for textured images. *Technometrics*, 56(3):291–301, 2014.

[232] Sarah L Taylor, Idris A Eckley, and Matthew A Nunes. Multivariate locally stationary 2d wavelet processes with application to colour texture analysis. *Statistics and computing*, 27(4):1129–1143, 2017.

[233] George Themelis, Jung S. Yoo, and Vasilis Ntziachristos. Multispectral imaging using multiple-bandpass filters. *Optics letters*, 33(9):1023, 2008.

[234] Ka Man Tsang et al. Study of fabric texture effects on visual and imaging color difference evaluation. 2016.

[235] Mihran Tuceryan and Anil K Jain. Texture analysis. *Handbook of pattern recognition and computer vision*, pages 235–276, 1993.

[236] U3m 1.1 introduction. https://raw.githubusercontent.com/vizoogmbh/u3m/master/u3m1.1/U3M.pdf. accessed in August 2021.

[237] Michael Unser. Texture classification and segmentation using wavelet frames. *IEEE Transactions on image processing*, 4(11):1549–1560, 1995.

[238] Philipp Urban, Tejas Madan Tanksale, Alan Brunton, Bui Minh Vu, and Shigeki Nakauchi. Redefining a in rgba: towards a standard for graphical 3d printing. *ACM Transactions on Graphics (TOG)*, 38(3):1–14, 2019.

[239] Muhammad Uzair, Arif Mahmood, and Ajmal Mian. Hyperspectral face recognition with spatiospectral information fusion and pls regression. *IEEE Transactions on Image Processing*, 24(3):1127–1137, 2015.

[240] C. Vanderriest. A fiber-optics dissector for spectroscopy of nebulosities around quasars and similar objects. *Publications of the Astronomical Society of the Pacific*, 92(550):858, 1980.

[241] Tibert Verhagen, Bart Van Den Hooff, and Selmar Meents. Toward a better use of the semantic differential in is research: An integrative framework of suggested action. *Journal of the Association for Information Systems*, 16(2):1, 2015.

[242] Poorvi L Vora and H Joel Trussell. Measure of goodness of a set of color-scanning filters. *JOSA A*, 10(7):1499–1508, 1993.

[243] Theodore V. Vorburger, Richard Silver, Rainer Brodmann, Boris Brodmann, and Jörg Seewig. Light scattering methods. In *Optical Measurement of Surface Topography*, pages 287–318. Springer, 2011.

[244] L. Weitzel, A. Krabbe, H. Kroker, N. Thatte, L. E. Tacconi-Garman, M. Cameron, and R. Genzel. 3d: The next generation near-infrared imaging spectrometer. *Astronomy and Astrophysics Supplement Series*, 119(3):531–546, 1996.

[245] Ning Xie, Gabrielle Ras, Marcel van Gerven, and Derek Doran. Explainable deep learning: A field guide for the uninitiated. *arXiv preprint arXiv:2004.14545*, 2020.

[246] xtex page on vizoo's website. https://www.vizoo3d.com/xtex-bundle. accessed in June 2021.

[247] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM transactions on graphics (TOG)*, 31(6):1–10, 2012.

[248] Yong Xu, Sibin Huang, Hui Ji, and Cornelia Fermüller. Scale-space texture description on sift-like textons. *Computer Vision and Image Understanding*, 116(9):999–1013, 2012.

[249] XB Yang and XB Huang. Evaluating fabric wrinkle degree with a photometric stereo method. *Textile Research Journal*, 73(5):451–454, 2003.

[250] Yong Yang, Mei Yang, Shuying Huang, Yue Que, Min Ding, and Jun Sun. Multifocus image fusion based on extreme learning machine and human visual system. *IEEE access*, 5:6989–7000, 2017.

[251] Juha Ylioinas, Abdenour Hadid, Yimo Guo, and Matti Pietikäinen. Efficient image appearance description using dense sampling based local binary patterns. In *Asian conference on computer vision*, pages 375–388. Springer, 2012.

[252] Hang Zhang, Jia Xue, and Kristin Dana. Deep ten: Texture encoding network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 708–717, 2017.

[253] Feng Zhou, Ju Fu Feng, and Qing Yun Shi. Texture feature based on local fourier transform. In *Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205)*, volume 2, pages 610–613. IEEE, 2001.

[254] Song Chun Zhu, Yingnian Wu, and David Mumford. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *International Journal of Computer Vision*, 27(2):107–126, 1998.

[255] Jana Zujovic, Thrasyvoulos N Pappas, and David L Neuhoff. Structural texture similarity metrics for image analysis and retrieval. *IEEE Transactions on Image Processing*, 22(7):2545–2558, 2013.

[256] Jana Zujovic, Thrasyvoulos N Pappas, David L Neuhoff, Rene Van Egmond, and Huib De Ridder. Subjective and objective texture similarity for image compression. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1369–1372. IEEE, 2012.

Article A

Michele Conni, Helene Midtfjord, Peter Nussbaum, and Phil Green. Dependence of texture classification accuracy on spectral information. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

Dependence of texture classification accuracy on spectral information

Michele Conni
Barbieri Electronic
Brixen, Italy;
Department of Computer Science
NTNU
Gjøvik, Norway
michele.conni@barbierielectronic.com

Helene Midtfjord
Department of Computer Science
NTNU
Gjøvik, Norway
helene.midtfjord@ntnu.no

Peter Nussbaum
Department of Computer Science
NTNU
Gjøvik, Norway
peter.nussbaum@ntnu.no

Phil Green
Department of Computer Science
NTNU
Gjøvik, Norway
philip.green@ntnu.no

***Abstract*—The influence of the number of spectral channels of spatial information on the accuracy of texture classification was evaluated by modelling the spectral sensitivity functions of a group of ideal imaging systems with a set of gaussian functions and then applying them to a set of hyperspectral images in order to simulate the response of each of their colour channels. Feature extraction and classification with different techniques were applied to the simulated data to assess the performance. It was shown that a significant enhancement of the accuracy was achieved, with a dependency on the approach adopted; it is therefore possible to estimate the optimal number of spectral channels for each method.**

***Index Terms*—colour texture classification, spectral imaging, feature extraction**

I. Introduction

Texture classification is one of the fundamental applications of texture analysis, as it can be exploited in any field in which high-level tasks, such as human-like perception and visual recognition of surfaces, are linked to data obtained with a digital camera. This type of analysis has became fundamental in multiple fields, e.g. food science [29], medicine [23], face recognition [9] and defect detection [14]. The camera-based approach can be expanded with additional input information, such as a surface topographical pattern of the sample [4] or multi-illumination analysis [11]; however, the accuracy improvement achieved with these is usually at the cost of increased complexity of the acquisition system.

Most classification approaches involve the use of *features* (descriptors): these elements, which are outputs of an initial computational step, consist of sets of parameters distinctive of the surface under study, and are derived from the available data with a variety of different techniques. The use of descriptors permits the reduction of dimensionality of the characterization problem, as an array of features normally has a dimension of tens or hundreds, which is considerably less data than an entire image. This reduction of dimensionality helps reduce the number of degrees of freedom for the characterization process. A clear example of the latter can be found in the Local Binary Pattern (LBP) operator (one of the most effective feature extraction models), which is robust to illumination changes in the texture images [7]. Usually, texture features are extracted from grey-level images. However, it has previously been shown that the addition of multispectral information enhances the performance of such analysis, both for colour [3] and spectral [13] data.

This paper aims to evaluate the dependence of classification accuracy on the number of spectral channels of the detector, thus obtaining indicators on the optimal measurement setup for texture classification. An increase of the amount of information is generally associated with a better definition of the classes; however, since this increase in input data corresponds to the enlargement of the data space through the addition of dimensions and a consequent increase in complexity [1]. This work is composed of the following steps: first, a set of hyperspectral images, consistent with the already existing texture databases, has been acquired; subsequently, a group of images, with a number of channels m increasing from 1 to n, has been simulated for each image by applying a set of m theoretical sensitivity functions. A selection of methods for feature extraction (grey-level co-occurrences matrices (GLCM), local binary patterns (LBP), Gabor transformation and wavelet transformation) has been then applied to the simulated dataset. Finally, the features extracted have been subjected to multiple classification algorithms, in order to obtain a value for the accuracy for different number of channels.

II. Data Acquisition

The data used in this paper are a collection of bidimensional spectral images. Three examples of pictures are illustrated in Fig. 1. The acquired databases consist of 191 sample images,

2018 Colour and Visual Computing Symposium (CVCS)
978-1-5386-5645-7/18/$31.0 © 2018 IEEE

(a) *Textile*. (b) *Leather*. (c) *Wood*.

Fig. 1: Illustration of three of the images in our database. The RGB images are simulated from the multispectral images with GLIMPS imaging spectroscopy image analysis.

divided into the six texture macro classes: textiles (155), leather (3), paper (8), plastic (13), wood (8), and metal (4). Textiles are again divided into nine micro classes: viscose (1), jute (1), polyamide (1), wool (2), mixed natural (4), linen (3), polyester (42), Trevira CS (48) and mixed synthetic (53). While each subclass contains different typologies of spatial texture distributions, additional subgroups comprising different colour gradations of the same visual pattern can be further defined. Some of the materials are translucent, and an opaque sheet of white paper (without optical brighteners) has been used as sample backing for each acquisition.

The hyperspectral images have been acquired with a HySpex VNIR-1800 [10], which is a linear CMOS detector with a spectral range $\Lambda \in [400, 1000]\,nm$ and a resolution $\Delta\lambda = 3.26\,nm$, i.e. 184 spectral channels. The setup is equipped with two 150W tungsten halogen light sources positioned at $\pm 45°$ with respect to the detector and a movable sample-holder table. The focal length of the optical system has been fixed to $30\,cm$, while the linear field of view (FOV) is approximately $10\,cm$. The dimensions of the samples are greater than or equal to $5\,cm \times 5\,cm$, thus allowing the extraction of patches of dimension 640×640 pixels from the multispectral images. These dimensions have been chosen to match the images of the already existing and widely used SpecTex database [25].

Being the raw output of the measurement proportional to the radiance detected, the data obtained has been processed to extract reflectance of the samples. Since the integration time could potentially vary from image to image, a reference grey patch with a known spectral reflectance has been inserted in each image; this has been used to determine the spectral power distribution of the light source in the image. The captured radiance has been divided by this light source spectral power distribution in order to get the desired reflectance $R(\lambda)$ of the object under analysis.

III. Methodology

As described in Section I, the workflow for the classification experiment can be subdivided into three main steps: sensitivity function filtering, feature extraction and actual texture classification. Here we give a brief overview of the techniques adopted for these tasks.

A. Filtering

The input to the process is the set of hyperspectral images described in section II, each element of which consists of an image with 184 spectral channels distributed on a $\Lambda \in [400, 1000]\,nm$ domain, corresponding to the visible and near infrared regions of wavelengths. The first step is the sensitivity function filtering, in which the images are filtered with ideal sensitivity functions of n different detectors, each one with an increasing number of channels m that goes from 1 to n. Thanks to this process, the acquisition of the same image from each of these devices can be simulated. For the definition of the spectral sensitivity functions, we referred to the model of a generic one-shot spectral filter array camera. In this case, the signal acquired is $\rho_i = \int_\Lambda R(\lambda) l(\lambda) S(\lambda) t_i(\lambda) d\lambda$ for the i^{th} spectral channel; in this expression, $R(\lambda)$ is the reflectance of the sample, $l(\lambda)$ is the spectral power distribution of the illumination, $S(\lambda)$ is the sensor spectral response and $t_i(\lambda)$ is the spectral transmission of the i^{th} filter. The response $S(\lambda)$ varies from camera to camera and depends on its sensor material structure. An example is illustrated in [24]. Spectral filters responses $t_i(\lambda)$, as assessed in [12], most often follow a Fabry-Pérot filter model, which for our purposes is replaceable with a Gaussian one with negligible error. In the workflow adopted, the cumulative quantity $T_i(\lambda) = S(\lambda) \cdot t_i(\lambda)$ has been taken into account for each channel i, and, under the assumption of a constant sensor response $S(\lambda)$, a Gaussian model has been adopted. Despite this is a simplification, it is considered suitable fo the analysis performed [12]. The effective sensitivity functions $T_i(\lambda)$, hence, have been defined as a collection of equi-gaussians with unitary maxima and variances chosen so that the union of the wavelength domain covered by their standard deviation corresponds to the complete bandwidth of the acquisition sensor ($\sum_{i=1}^{m} 2\sigma_i = \Lambda$). This approach results in a high level of inter-channel correlation, but this ensures that stimuli from all spectral regions are represented.

B. Feature extraction

Having obtained for each picture a set of n images, the m^{th} of which is characterized by m colour channels, the feature extraction is applied to them. In [16], Palm sorted the possible approaches to extract features from a colour image into three groups: parallel, sequential and integrative. The parallel method involves splitting each picture into a grey-scale component and a set of colour histograms, before extracting features from these two elements. The sequential partitions the image based on the clusters of the colour histogram, while the integrative method involves treating each colour channel as a grey-scale image. The integrative technique is adopted in our work, since they permit us to analyze each colour channel separately and to compare features extracted from images with different number of channels. Their simplest implementation (referred to as *single-channel*) consists of applying grey-scale texture analysis to each distinct channel, while more complex ones (*multi-channel*) also quantify the correlation between channels. We treat each simulated colour channel as a grey-scale image, because the number of permutations between

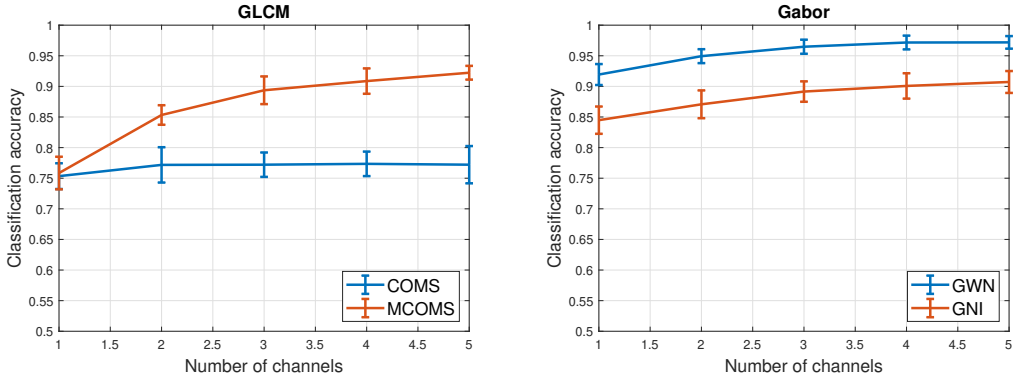


Fig. 2: Accuracy plotting for 3-Nearest Neighbour classifier, 5 channels and GLCM and Gabor feature vectors.

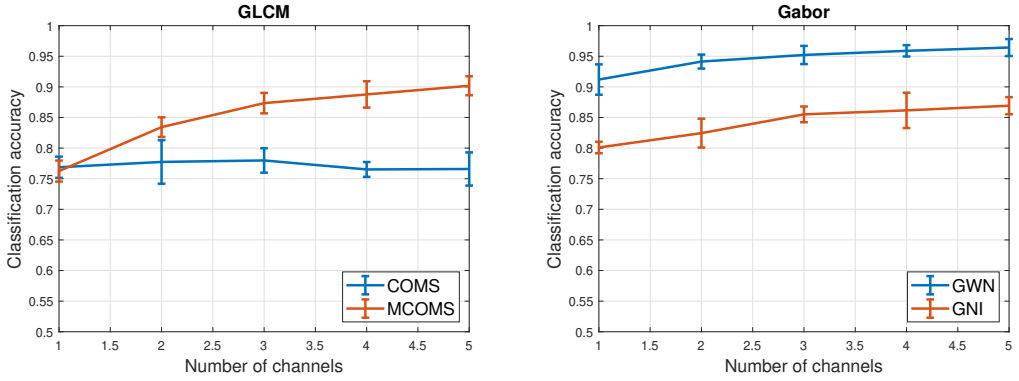


Fig. 3: Accuracy plotting for Random Forest classifier, 5 channels and GLCM and Gabor feature vectors.

channels required by the multi-channel approach is factorial in the number of elements, and requires significant computational time for large numbers of channels.

Four feature extraction techniques have been implemented: GLCM, LBP, Gabor transform and wavelet transform. Each of these has been used to extract a set of 14 vectors with different dimensions, and the classification has been performed on each vector separately; this selection has been derived based on the results of [3]. For GLCM [8], the calculation was performed at scales 1 to 8 and at 8 angles (equi-spaced by 45°), the average at each scale is computed, and contrast, homogeneity and energy are then calculated; from this, two vectors are derived, one containing only the results for scale 1 ($CMOS$, 3 features) and another containing all of them ($MCMOS$, 24 features). For LBP, 4 vectors are computed, 2 with uniform patterns [15] and 2 with the Completed Local Binary Count (CLBC) [28]: from uniform patterns, a 10 feature vector accounting for 8 neighbouring pixels at radius 1 (LBP^{riu}) and a 54 feature one that concatenates the previous one with the vector obtained with the same technique at radius 2 and 16 neighbours and the one at radius 3 and 24 neighbours, for a total of 54 elements ($MLBP^{riu}$). For Local Binary Count, both vectors are obtained by applying the count to the structures used for the uniform pattern, thus getting to a 9 ($CLBC$) and 53 ($MCLBC$) features respectively. Two feature vectors were derived with Gabor filters [19], one normalized for illumination invariance (GNI) and the other without normalization (GWN). They are formed by 48 parameters each, corresponding to the mean and the standard deviation of the magnitudes of the Gabor transformed image. The transformation has been obtained with filters characterized by different number of frequencies ($\in 1, ..., n_F$) and orientations ($\in 1, ..., n_O$). The filtering parameters have been chosen taking the results of [2] into account: $F_M = 0.327$, $n_F = 4$, $F_r = \sqrt{2}$, $n_O = 6$ and $\eta = \gamma = 0.5$. Finally, six wavelet transform vectors were used, two of which are generated with

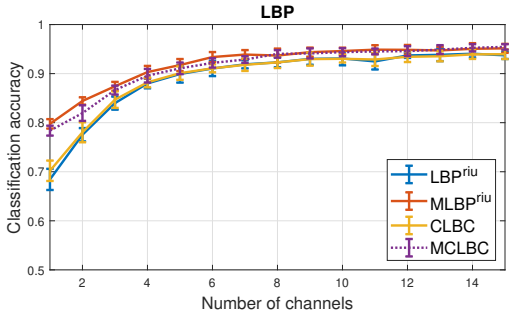


Fig. 4: Accuracy plotting for 3-Nearest Neighbour classifier, 15 channels, LBP and wavelet feature vectors.

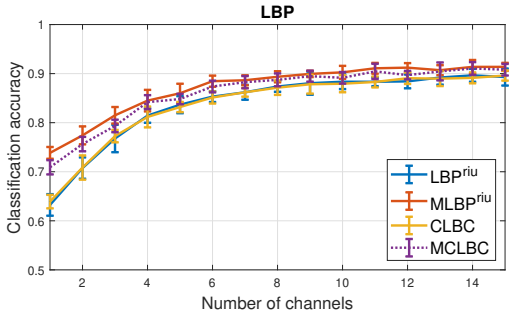


Fig. 5: Accuracy plotting for Random Forest classifier, 15 channels, LBP and wavelet feature vectors.

Haar filters, two with the Daubechies ones and two with the dualt-tree complex wavelet transform (DTCWT, [21]). Haar and Daubechies vectors contain mean and variance of the energy of the wavelet transformed image; both of them have 3 levels of decomposition, while the size of the Daubechies filter is 4. Two vectors have been derived from each filter: one containing the mean values of 10 transformations ($HaarME$ and $DaubechiesME$), the other which adds the corresponding variances ($Haar$ and $Daubechies$). The two vectors obtained with the DTCWT are composed respectively by mean and variance ($DTCWT$), and variance and entropy statistics ($DTCWT-VH$); overall, 3 scales and 6 directions are taken into account for the filters, for a total of 36 elements for each vector. The GLCM, LBP and Gabor filtering functions of the `skimage.feature` Python library [22] were used to perform the implementation, together with PyWavelets [18] and `dtcwt` [5] modules.

C. Classification

In the third step, the extracted feature vectors are taken as inputs for classification algorithms; in line with the texture classification literature [3], the feature extraction described in Section III-B has been performed on patches of 160×160 pixels of the samples. This means that 16 patches have been extracted from each image, and the image number of the original picture has been used to define the class of each patch. Given the limited size of the dataset (~ 3000 images), the subdivision into training and testing was 70%-30%, respectively. Two classifiers have been selected following the results of [6]: the 3-NN (k-Nearest Neighbours with $k = 3$) algorithm is one of the most basic approaches, while the Random Forest (RF) classifier has shown to generally provide precise solutions. Python's machine-learning library `scikit.learn` [17] was used; it comprises a large variety of classifiers, 3-NN and RF between the others. After features have been extracted for the whole dataset, and the training step has been performed on the 70% training sub-set of the images, the accuracy of the application of the trained model on the test set (composed by the remaining 30%) is evaluated and taken as final output. Since some of the images showed some defect, only a fraction of the dataset (texture inhomogeneities), consisting of 170 pictures, has been used for the classification.

IV. Results and Conclusions

A preliminary experiment has been conducted, in which the procedure described was implemented to classify the SpecTex texture database [25]. This has been found to be effective, showing however a strong variation with respect to

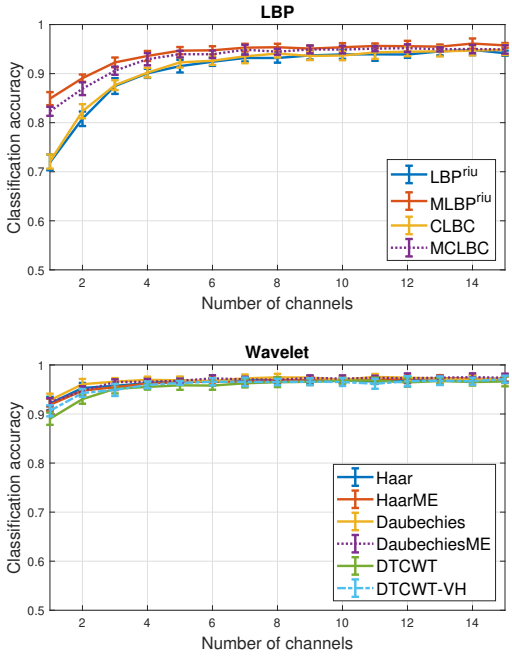


Fig. 6: Accuracy plotting for 3 Nearest Neighbour classifier, 15 channels, LBP and wavelet feature vectors; visible spectral domain.

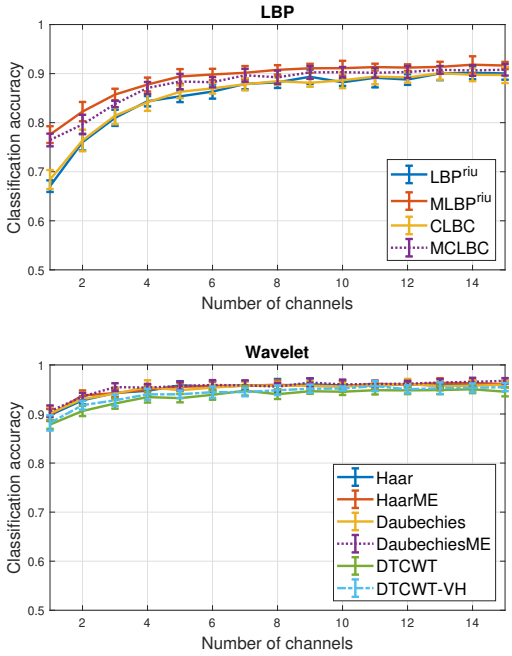


Fig. 7: Accuracy plotting for Random Forest classifier, 15 channels, LBP and wavelet feature vectors; visible spectral domain.

the expected trend for images with fewer colour channels; this variation has been subsequently linked to the use of a feature selector method.

Fig. 2 and Fig. 3 show the percentages of classification accuracy obtained with Gabor and GLCM features and a simulation of up to 5 channels, with both 3-NN and RF classifiers. More extended results are shown in Fig. 4 and Fig. 5 (up to 15 channels) for LBP and wavelet transform; due to the large amount of computational time demanded by them, GLCM and Gabor filter features extraction methods have been neglected in these cases. Nonetheless, the results for a number of channels up to 5 show the positive effect of the spectral information on these approaches. All these results have been derived from a 10 fold repetition of the classification, from which the average values have been extracted, as well as the 95% confidence intervals (for which a Student's t distribution model for the sample size has been adopted). It can be seen that, for the first five channels, there is in fact a monotonic increase of the gain, due to the larger number of samples taken into account. The total increase over the channels is on average $\sim 5\%$ for GLCM, Gabor and wavelet features and $\sim 10\%$ for LBP; the gain of each step, however, decreases progressively. It can be noted that the results of the two classifiers are almost identical for most of the approaches; there is however a clear difference of $\sim 5\%$ in the case of the Local Binary Pattern features, with the highest numbers belonging to the 3-NN classifier. The only technique that is not affected by the number of channels is the $COMS$, behavior that can be related to its simplicity. Moreover, it is clear that the difference between the distinct wavelet approaches is minimal.

In Fig. 4 and Fig. 5, the results with 15 channels show a trend of the classification performances analogous to the 5 channel one; a gradual reduction of the accuracy gradient, which we attribute to the distribution of the texture information in the spectral domain, is clearly visible. The performance of each approach has an asymptotic saturation value, which permits to derive an optimal number of channels over which the additional spectral information would be useless. A possible approach to derive this parameter is to fit the data with a non-linear function, such as a polynomial or the step response discussed in [26].

Finally, the effect of the bandwidth of acquisition has been taken into account: since the wavelength range under study is quite broad, and it is in practice quite costly to cover with a spectral sensor, we reproduced the 15 channels analysis of LBP- and wavelet-derived features on the visible domain ($\lambda \in$

$[400, 730]\, nm$). The results are shown in Fig. 6 and Fig. 7: a performance enhancement of $\sim 5\%$ is visible for the lower number of channels. This can be related to the lack of useful information in the infrared region, in which case the elements of each feature vector related to such a region would be similar for a great percentage of the samples, thus adding difficulty to the correct class sorting.

To conclude, this work derived the trend of the classification accuracy of a set of textiles depends on the number of available spectral channels; this dependence has a saturation value, which is related to the spectral sensitivity functions and the bandwidth of the detection setup, as well as to the feature extraction and classification methods. We propose that it is possible to calculate such a value for a given set of data, in order to select the best acquisition setup to classify them. In Figs. 4-7 it can also be observed that a number of channels larger than 3 does in general positively affect the accuracy of classification, while a number higher than 6 is not significantly beneficial.

There are many potential improvements to this analysis, the most straightforward one being to evaluate the performance of a wider range of feature extraction methods, here not taken into account, such as fractal-related ones [20] and machine learning ones [27], along with different classifiers [6]. Moreover, the dependence of the performance on the shape and the position of the sensitivity functions applied, as well as the effect of a more complex parallel feature extraction method, have not been investigated; as discussed, the Gaussian model exploited has some limitations. The application of sensitivity filters obtained directly from real multispectral cameras to the hyperspectral images could be used to evaluate the applicability of the model proposed. The data thereby acquired could also be exploited to evaluate how the camera chromatic calibration influences the classification accuracy.

ACKNOWLEDGMENT

The authors would like to thank Hilda Deborah, Haris Ahmad Khan and Jean-Baptiste Thomas for their precious suggestions and advice on the topics related to the paper.

REFERENCES

[1] Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001, January). On the surprising behavior of distance metrics in high dimensional space. In International conference on database theory (pp. 420-434). Springer, Berlin, Heidelberg.

[2] Bianconi, F., & Fernández, A. (2007). Evaluation of the effects of Gabor filter parameters on texture classification. Pattern Recognition, 40(12), 3325-3335.

[3] Cernadas, E., Fernández-Delgado, M., González-Rufino, E., & Carrión, P. (2017). Influence of normalization and color space to color texture classification. Pattern Recognition, 61, 120-138.

[4] DeGol, J., Golparvar-Fard, M., & Hoiem, D. (2016). Geometry-informed material recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1554-1562).

[5] dtcwt module documentation: https://pypi.org/project/dtcwt/.

[6] Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems. J. Mach. Learn. Res, 15(1), 3133-3181.

[7] Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. IEEE Transactions on Image Processing, 19(6), 1657-1663.

[8] Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. IEEE Transactions on systems, man, and cybernetics, (6), 610-621.

[9] Huang, D., Shan, C., Ardabilian, M., Wang, Y., & Chen, L. (2011). Local binary patterns and its application to facial image analysis: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), 765-781.

[10] HySpex VNIR-1800 features: https://www.hyspex.no/products/vnir_1800.php.

[11] Kampouris, C., Zafeiriou, S., Ghosh, A., & Malassiotis, S. (2016, October). Fine-grained material classification using micro-geometry and reflectance. In European Conference on Computer Vision (pp. 778-792). Springer, Cham.

[12] Lapray, P. J., Thomas, J. B., Gouton, P., & Ruichek, Y. (2017). Energy balance in Spectral Filter Array camera design. Journal of the European Optical Society-Rapid Publications, 13(1), 1.

[13] Liu, X., & Wang, D. (2003). Texture classification using spectral histograms. IEEE transactions on image processing, 12(6), 661-670.

[14] Ngan, H. Y., Pang, G. K., & Yung, N. H. (2011). Automated fabric defect detection—a review. Image and Vision Computing, 29(7), 442-458.

[15] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7), 971-987.

[16] Palm, C. (2004). Color texture classification by integrative co-occurrence matrices. Pattern recognition, 37(5), 965-976.

[17] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of machine learning research, 12(Oct), 2825-2830.

[18] PyWavelets module documentation: http://pywavelets.readthedocs.io.

[19] Randen, T., & Husoy, J. H. (1999). Filtering for texture classification: A comparative study. IEEE Transactions on pattern analysis and machine intelligence, 21(4), 291-310.

[20] Sarkar, N., & Chaudhuri, B. B. (1992). An efficient approach to estimate fractal dimension of textural images. Pattern recognition, 25(9), 1035-1041.

[21] Selesnick, I. W., Baraniuk, R. G., & Kingsbury, N. C. (2005). The dual-tree complex wavelet transform. IEEE signal processing magazine, 22(6), 123-151.

[22] Scikit-image feature extration documentation: http://scikit-image.org/docs/dev/api/skimage.feature.html.

[23] Song, Y., Cai, W., Zhou, Y., & Feng, D. D. (2013). Feature-based image patch approximation for lung tissue classification. IEEE transactions on medical imaging, 32(4), 797-808.

[24] Sony Pregius Global Shutter CMOS Imaging Performance. Available online: http://www.ptgrey.com/support/downloads/10414

[25] SpecTex - HyperSpectral Texture Image Database: http://www.uef.fi/web/spectral/spectex.

[26] Tay, T. T., Mareels, I., & Moore, J. B. (2012). High performance control. Springer Science & Business Media.

[27] Varma, M., & Zisserman, A. (2009). A statistical approach to material classification using image patch exemplars. IEEE transactions on pattern analysis and machine intelligence, 31(11), 2032-2047.

[28] Zhao, Y., Huang, D. S., & Jia, W. (2012). Completed local binary count for rotation invariant texture classification. IEEE transactions on image processing, 21(10), 4492-4497.

[29] Zheng, C., Sun, D. W., & Zheng, L. (2006). Recent applications of image texture for evaluation of food qualities—a review. Trends in Food Science & Technology, 17(3), 113-128.

Article B

Michele Conni, Peter Nussbaum, and Phil Green. The effect of camera calibration on multichannel texture classification. *Journal of Imaging Science and Technology*, 65(1):10503–1, 2021

This paper is not included due to copyright restrictions

Article C

Michele Conni and Hilda Deborah. Texture stationarity evaluation with local wavelet spectrum. In *London Imaging Meeting*, volume 2020, pages 24–27. Society for Imaging Science and Technology, 2020

Article D

Michele Conni, Hilda Deborah, Peter Nussbaum, and Phil Green. Visual and data stationarity of texture images. *Journal of Electronic Imaging*, 30(4):043001, 2021

Visual and Data Stationarity of Texture Images

Michele Conni[a, b], **Hilda Deborah**[a], **Peter Nussbaum**[a], **Phil Green**[a]
[a]Department of Computer Science, Norwegian University of Science and Technology (NTNU); Gjøvik, Norway
[b]Barbieri Electronic; Brixen, Italy

Abstract. The stationarity of a texture can be considered a fundamental property of images, although the property of stationarity is difficult to define precisely. We propose a stationarity test based on multiscale, locally stationary, 2D wavelets. Three separate experiments were performed to evaluate the capabilities and the limitations of this test. The experiments comprised a chessboard stationarity analysis, two classification tasks, and a psychophysical experiment. The classification tasks were performed on 110 texture images from a texture database; in one sub-task five texture feature vectors were extracted from each image and the classification accuracy of two classical methods compared, while in the second sub-task the classification accuracy of several methods was compared to the descriptors defined for each image within the database. In the psychophysical experiment the correlation between the classification results and observer judgements of texture similarity were determined. It was found that a combination of wavelet shrinkage and rotation-invariant local binary pattern (LBP) best predicted the observer response. The results show that the proposed stationarity test is able to provide relevant information for texture analysis.

Keywords: texture stationarity, visual texture, texture classification.

1 Introduction

A one-dimensional temporal signal is said to be *stationary* if its local statistical properties are constant in time.[1] The same concept can be extended to the field of texture analysis, where it is widely used. It is, in fact, a fundamental assumption of global texture models, such as Markov random fields, auto-correlation functions and the well-known Tamura features.[2] Consequently, if the texture is not stationary, these techniques give a flawed representation of the signal. In fact, this fails to depict the actual mathematical properties of the texture, because it extracts an average behaviour that neglects any change of local features in the image. For such a case, a texture can be divided into stationary sub-regions with a segmentation algorithm which autonomously partitions an image into multiple homogeneous areas.[2] However, this increases the computational complexity of the analysis and, as discussed in Ref. 3, prior knowledge of the stationarity of the sample would still be needed.

Currently, the stationarity property of a texture image has a dual meaning. From a mathematical point of view, the term *stationarity* means that the average of the data generating process, which gives rise to the image observed, is the same everywhere in the image,[3] and that its distribution is essentially regular, i.e., its variance is finite and its covariance is dependent only on the distance between pixels. We use the term *data stationarity* to refer to this definition. On the other hand, in Ref. [4, pg.80], Petrou and Sevilla state that "*a stationary texture image is an image which contains a single type of texture*". This suggests an interpretation of stationarity which is more related to the human visual system. We name this second definition *visual stationarity*. This interpretation is more complex than the first one, as it touches the border of linguistics, i.e., the understanding of the concept of *a single type of texture*, which would probably depend on the context of analysis. This situation is not uncommon, for example even the similar but more widely used term *homogeneous* has a fuzzy meaning in the research of human perception of texture.[5]

An example that shows the practical need for a shared definition of data and visual stationarity is suggested by Ref. 6. In this article, the authors classified a wide variety of existing texture databases according to certain characteristics. Among these characteristics is texture stationarity, with explicit mention to the definition given in Ref. 3. In Appendix C we show that this partition is inconsistent with the results of the stationarity test employed in the current paper (see Sec. 2). This mismatch clearly highlights the discrepancy between data and visual stationarity.

In a previous study,[7] we analysed the concept of *data stationarity* and we expanded a framework developed to evaluate image stationarity[8] to account for multiple scales. The choice of scale has proven to be of great importance for texture analysis.[9] The link between this mathematical approach and *visual stationarity* is discussed in Ref. 8, in which the stationarity test was applied to images of pilled fabric, and the results subsequently compared to the authors' evaluation of the visual stationarity of the images. This however provides limited experimental psychophysical data on which to base firm conclusions. In this work, we wish to remedy this gap in the literature. To this end, we provide an investigation of the relationship between visual and data stationarity, using images of our own dataset as well as from a subset of the Describable Texture Dataset (DTD).[10] The choice of DTD is based on its texture categorization and annotation by human observers, thus incorporating the complexity of human perception of texture. Additionally, data stationarity analyses of an alternative texture database can be found in Appendix C.

2 Texture Stationarity

In the field of visual texture analysis, the conjecture proposed by Julesz, stating that "*whereas textures that differ in their first- and second-order statistics can be discriminated from each other, those that differ in their third- or higher-order statistics usually cannot*,"[11] is a good approximation of how human perception works. Many texture feature extraction techniques therefore assume that their image targets are second-order stationary,[3,12] i.e., the process generating these images has a constant mean, a finite variance, and a covariance that is a function of pixel distance $cov(X_{\mathbf{r}_1}, X_{\mathbf{r}_2}) = \gamma(\mathbf{r}_1 - \mathbf{r}_2)$. In Ref. 8, Taylor and colleagues employed these premises to develop an image stationarity test for a single realization of a generating statistical process. The test interprets each image as a Locally Stationary Two-Dimensional Wavelet (LS2W) process, and it evaluates the constancy of its power spectrum to estimate its stationarity. We introduce and describe it in Section 2.1, and we propose a variation to it in Section 2.2. Frequently used mathematical notations are also summarized in Table 1.

2.1 The LS2W stationarity test

A *mother wavelet* $\psi(x)$ is a compact support function with oscillatory characteristics,[13–15] $x \in \mathbb{R}$, which, together with an auxiliary function $\phi(x)$ called a *father wavelet*, can be used to form a complete functional basis on $L^2(\mathbb{R})$. This functional basis $\{\psi_{j,k}, \phi_{j,k}\}_{j,k\in\mathbb{Z}}$ is achieved by scaling and shifting $\psi(x)$ and $\phi(x)$, with j and k indicating the scaling and shifting indices respectively. On one hand, the shifting gives the possibility of representing local segments of the signal, while on the other the scaling allows it to represent the fine or coarse structures contained therein. As discussed in Ref. 16, a discrete version of such a basis function can be obtained by associating two compactly supported mother and father wavelets ψ and ϕ with a suitable pair of low-pass and high-pass filters $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$. In this case, a discrete wavelet at scale j is a vector $\psi_j = (\psi_{j,0}, \ldots, \psi_{j,N_j-1})$, where $N_j = (2^j - 1)(N_h - 1) + 1$, $N_h \neq 0$, $\psi_{-1,n} = g_n$ and $\psi_{j-1,n} =$

Table 1: Frequently used mathematical notations.

$\psi(x), \phi(x)$	Mother and father wavelets, respectively
γ	Covariance function $\gamma : \mathbb{Z}^2 \to \mathbb{R}$ of a stationary image
$\mathbf{r}_i$	Location or coordinate of arbitrary pixel i, with $\mathbf{r} \in \mathbb{Z}^2$
j	Wavelet scaling index, where $j \in \mathbb{Z}^+$
$k, \mathbf{k}$	Wavelet shifting index and vector, respectively, $k \in \mathbb{Z}$ and $\mathbf{k} = (k_1, k_2),\ k_i \in \mathbb{Z}$
$\mathbf{u}$	Image coordinates after wavelet shifting, $\mathbf{u} = \mathbf{r} + \mathbf{k}$
l	Wavelet direction index, where $l \in \{H, V, D\}$
h, g	Discrete low- and high-pass filters, respectively
N_h	Number of nonzero elements in h, $N_h = \#(h) \neq 0$
N_j	Number of elements of the discrete wavelet ψ_j at scale j
R, C	Number of rows and columns in an image, respectively, expressed in terms of a power of 2, $R = 2^m, C = 2^n, n, m \in \mathbb{N}^+$
$\mathbf{R}$	Dimension of a greyscale image, $\mathbf{R} = (R, C)$
$w^l_{j,\mathbf{u}}$	Coefficient of the wavelet transform
$\xi^l_{j,\mathbf{u}}$	Zero-mean random orthonormal increment sequence
J	Lowest significant scale, $J(R, C) = \log_2\{\min(R, C)\}$
$X_{\mathbf{r};\mathbf{R}}$	Generic LS2W process with dimension $\mathbf{R}$
$\mathbf{z}$	Normalized spatial coordinate, $\mathbf{z} = \mathbf{u}/\mathbf{R} \coloneqq (u/R, v/C), \mathbf{z} \in (0, 1)^2$
S^l_j	Local wavelet spectrum
$d^l_{j,\mathbf{u}}$	Empirical mother wavelet coefficients
$\mathbf{I}(\mathbf{u})$	Local wavelet periodogram (LWP) as an estimator for S^l_j
A_J	LWP correction matrix
$\hat{\mathbf{S}}(\mathbf{u})$	Estimator for LWP, $\hat{\mathbf{S}}(\mathbf{u}) = A_J^{-1}\mathbf{I}(\mathbf{u})$, composed by the elements $\hat{S}^l_j(\mathbf{u})$
T_{ave}	Departure from constancy of an estimated LWP $\hat{\mathbf{S}}(\mathbf{u})$
B	Number of repetitions of the bootstrap loop
p	p-value of the stationarity test for $\hat{\mathbf{S}}$
$\eta(j, l)$	Index of scale-direction pair, $\eta \in \{1, \ldots, 3J\}$
$p_{\eta(j,l)}$	p-value of the stationarity test for $\hat{S}^l_j$
$\mathbf{p}$	Vector of $p_{\eta(j,l)}$ at various dyadic scales and directions, $\mathbf{p} = (p_{\eta=1}, \ldots, p_{\eta=3J})$
p_j	p-value of the stationarity test for $\hat{S}_j$
$\mathbf{p}_j$	Vector of p_j at various dyadic scales, $\mathbf{p}_j = (p_{j=1}, \ldots, p_{j=J})$
$X_{wn}(\mathbf{r})$	Bi-dimensional white-noise process
$N(\mu, \sigma)$	Normal distribution with mean μ and standard deviation σ

$\sum_k h_{n-2k}\psi_{j,k}$, $\forall n \in [0, \ldots, N_{j-1} - 1]$. Such a basis can be easily expanded in two dimensions and applied to images. This is achieved by substituting k with $\mathbf{k} = (k_1, k_2)$ and introducing a direction index $l \in \{H, V, D\}$. l is employed to mix both father and mother wavelets, to ensure the completeness of the basis. Its values are H for horizontal, V for vertical, and D for diagonal. The corresponding 2D fundamental wavelets are defined as $\psi^H_{j,\mathbf{k}} = \phi_{j,k_1}\psi_{j,k_2}$, $\psi^V_{j,\mathbf{k}} = \psi_{j,k_1}\phi_{j,k_2}$ and $\psi^D_{j,\mathbf{k}} = \psi_{j,k_1}\psi_{j,k_2}$. A generic discrete wavelet at scale j in a given decomposition direction l can

then be expressed as in Eq. (1).

$$\psi_j^l = \begin{bmatrix} \psi^l_{j,(0,0)} & \cdots & \psi^l_{j,(0,N_j-1)} \\ \vdots & \ddots & \vdots \\ \psi^l_{j,(N_j-1,0)} & \cdots & \psi^l_{j,(N_j-1,N_j-1)} \end{bmatrix}. \tag{1}$$

The family of wavelets $\{\psi_j^l\}$ derived from the definition of Eq. (1) was used in Ref. 17 to define a random field modeling framework, called the *locally stationary two-dimensional wavelet field* (LS2W). The idea is to apply the complete set of 2D discrete wavelet matrices as filters on an image to calculate its wavelet coefficients in various pixel positions, determined by the shift $\mathbf{k}$. The approach proposed in Ref. 17 exploits only dyadic scales, while the filters are applied on every possible position of the image. Mathematically, a generic image of dimensions $\mathbf{R} = (R, C)$ can be generated with an LS2W process as in Eq. (2), where $\{w^l_{j,\mathbf{u}}\}$ are the wavelet coefficients and $\{\psi^l_{j,\mathbf{u}}(\mathbf{r}) = \psi^l_{j,\mathbf{u}-\mathbf{r}}\}$ are 2D discrete non-decimated wavelets with orientation l, scale j and shifted coordinate $\mathbf{u}$. Each coefficient $w^l_{j,\mathbf{u}}$ quantifies how large the contribution of the corresponding wavelet $\psi^l_{j,\mathbf{u}}(\mathbf{r})$ is in defining the process. $\{\xi^l_{j,\mathbf{u}}\}$ is a zero-mean random orthonormal increment sequence which allows stochastic structure to be encapsulated in the model. The dependence on the image dimension $\mathbf{R}$ is included to make the link with the lowest significant scale $J(R, C) = \log_2\{\min(R, C)\}$ explicit. Further on, it will be considered as implicit.

$$X_{\mathbf{r};\mathbf{R}} = \sum_l \sum_{j=1}^{\infty} \sum_{\mathbf{u}} w^l_{j,\mathbf{u};\mathbf{R}} \psi^l_{j,\mathbf{u}}(\mathbf{r}) \xi^l_{j,\mathbf{u}} \tag{2}$$

The *local wavelet spectrum* (LWS) of an LS2W process $X_\mathbf{r}$ can be considered as a Power Spectral Density for the stationary wavelet transform $S^l_j(\mathbf{z}) \approx w^l_j(\mathbf{u}/\mathbf{R})^2$. Here, $\mathbf{z} \in (0, 1)^2$ is a normalized spatial coordinate $\mathbf{z} = \mathbf{u}/\mathbf{R} \coloneqq (u/R, v/C)$ and, for a stationary process, $S^l_j(\mathbf{z})$ is a constant function of $\mathbf{z}$ $\forall j, l$.[17] Therefore, an estimate of the LWS can be used to assess the stationarity of an image.[8]

In Ref. 17, Eckley and colleagues proposed the *local wavelet periodogram* (LWP) as an estimator for $S^l_j(\mathbf{z})$. It is expressed as in Eq. (3), where $d^l_{j,\mathbf{u}}$ are the empirical mother wavelet coefficients of the image. The fact that the father wavelet coefficients are not included in Eq. (3) implies the independence of the LWP from the mean value of the process under study. This estimator is biased, but it can be corrected by multiplying it with the inverse of the two-dimensional discrete autocorrelation wavelet matrix A_J, obtaining $\hat{\mathbf{S}}(\mathbf{u}) = A_J^{-1}\mathbf{I}(\mathbf{u})$.[18] Prior to the correction, wavelet shrinkage has also been applied to each level of LWP to increase the consistency of the estimator.[17] $\hat{\mathbf{S}}(\mathbf{u})$ is an array with four dimensions, i.e., two for the spatial coordinates $\mathbf{u}$, one for scale j, and one for direction l, to reach a total of $R \times C \times J \times 3$ elements.

$$\mathbf{I}(\mathbf{u}) = \{I^l_{j,\mathbf{u}}\} = \{|d^l_{j,\mathbf{u}}|^2\} = \left\{ \left(\sum_{\mathbf{r}} X_\mathbf{r} \psi^l_{j,\mathbf{u}}(\mathbf{r}) \right)^2 \right\} \tag{3}$$

The stationarity test introduced in Ref. 8 employs as test statistic a departure from constancy $T_{ave}\{\hat{\mathbf{S}}\}$ in Eq. (4), which is the variance of the values of $\hat{\mathbf{S}}(\mathbf{u})$, averaged over scales j and direc-

tions l.

$$T_{ave}\{\hat{\mathbf{S}}\} = (3J)^{-1} \sum_{l} \sum_{j=1}^{J} \mathrm{var}_{\mathbf{u}}(\hat{\mathbf{S}}(\mathbf{u})) \tag{4}$$

Since the original distribution is unknown *a priori*, i.e., the algorithm operates on a single realization of the LS2W process $X_{\mathbf{r}}$, it is simulated with a bootstrap operation to infer its characteristics from the input image. Then, the p-value of the stationarity test is calculated by comparing the T_{ave} of the various bootstrap iterations with that of the original image. Mathematically, this can be expressed as $p = \frac{1+\#\{T_{ave}^{obs} \leq T_{ave}^{(i)}\}}{B+1}$, with obs indicating the observed image, index i specifying the various bootstrap instantiations, and B the total number of repetitions of the bootstrap loop.

2.2 Proposed approach to texture stationarity

As discussed in Ref. 8, it is also possible to test each scale-direction spectral plane for constancy. This is achieved by defining a test statistic $T_{\eta(j,l)}$ as shown in Eq. (5).

$$T_{\eta(j,l)}\{\hat{\mathbf{S}}\} := T\{\hat{S}_j^l\} = \mathrm{var}_{\mathbf{u}}(\hat{S}_j^l(\mathbf{u})) \tag{5}$$

With these test statistics it is possible to perform a stationarity test at every scale j and direction l. For each of these tests, a p-value $p_{\eta(j,l)} = \frac{1+\#\{T_{\eta(j,l)}^{obs} \leq T_{\eta(j,l)}^{(i)}\}}{B+1}$ can be defined. The $p_{\eta(j,l)}$s can be grouped into a vector $\mathbf{p}$ to understand the degree of stationarity of an image at dyadic scales $2^j,\ \forall j \in \mathbb{Z}^+, j < J(R,C)$ and for direction $l \in \{H, V, D\}$. Note that the calculation of p is non-linear in T_{ave}, which means that the average value of the vector $\mathbf{p}$ is different from the p-value of the image ($\overline{p_{\eta(j,l)}} \neq p$). Given that the family of wavelets $\{\psi_j^l\}$ is composed by orthogonal filters, each test is independent from the others. Such an approach is similar to the $\texttt{Bootstat}^{\mathrm{mh}}{}_{\texttt{LS2W}}$ framework introduced in Ref. 8, which however is used to probe the stationarity of the whole image, and not scale by scale. To achieve that, the $\texttt{Bootstat}^{\mathrm{mh}}{}_{\texttt{LS2W}}$ applies a multiple hypothesis testing scenario discussed in Ref. 19. In our case it is not necessary to resort to this correction method since we define $3J$ distinct hypothesis tests, one for each η.

Finally, we can define a set of p-values under the null hypothesis of stationarity of scale j. We can define these as $p_j = \frac{1+\#\{T_j^{obs} \leq T_j^{(i)}\}}{B+1}$, $T_j\{\hat{\mathbf{S}}\} := \frac{T_{\eta(j,H)}\{\hat{\mathbf{S}}\}+T_{\eta(j,V)}\{\hat{\mathbf{S}}\}+T_{\eta(j,D)}\{\hat{\mathbf{S}}\}}{3}$. This corresponds to Eq. (8) of Ref. 8, averaged only over the wavelet directions. We gather these values in the vector $\mathbf{p}_j$.

p $\mathbf{p}$ and $\mathbf{p}_j$ are all results of statistical tests, under the null hypothesis of stationarity. Therefore, a threshold α of the test significance level can be chosen. In the present paper, we used Haar wavelets and we set $B = 100$ and $\alpha = 5\%$, where not stated otherwise.

3 Chessboard Stationarity

The relevance of LS2W and the related stationarity test in computer vision applications has been discussed in various articles.[8,17,20,21] However, to our knowledge, its relationship to the human perception process and appearance analysis has not been previously addressed. A simple way to obtain some initial insights is by extracting the values p and $\mathbf{p}$ from images that have a regular structure. Our goal is therefore to investigate whether the tests introduced in the previous Section define these images as data stationary.

We applied the stationarity test on an 8-bit 128×128 chessboard image (Fig. 1a) with binary values (0 on black patches, 255 on white ones). Following the classification discussed in Ref. 3, this is a regular periodic pattern with a black square primitive, which thus can be analysed with a shape grammar. According to the definition in chapter four of the same Reference, the image can be perceived as filled with a single texture, i.e. it is visually stationary. Since the LS2W model assumes that the image analysed has a stochastic structure, we added a degree of random noise to the pixels of the chessboard picture: a two-dimensional white Gaussian noise process $X_{wn}(\mathbf{r}) \sim N(0, \sigma_{wn})$, with $\sigma_{wn} = 10 \cdot n,\ \forall n \in [0, 10]$. Only moments of order two and higher of the original checkerboard picture are influenced by the procedure because the distribution has an average value of zero. Given that for all of these images the p-value of the LS2W test is $p = 1$, they are data stationary according to the test. In this case, the scale analysis does not add any additional information, because $\mathbf{p}$ is constant and unitary $\mathbf{p} = \{p_{\eta(j,l)} = 1,\ \forall j \in \mathbb{Z}^+, \forall l \in \{H, V, D\}\}$. Interestingly, the results of the stationarity tests are the same for both the stochastic ($\sigma_{wn} \neq 0$) and the deterministic ($\sigma_{wn} = 0$) patterns that were analysed. This is probably due to the fact that the added noise is second-order stationary, and thus does not influence the result of the calculation, which is mainly dictated by the deterministic base. The zero mean of the overall image, which is assumed by the LS2W methodology, is, as discussed, independent on the noise distribution. This property is further ensured by the definition of the LWP itself (see Eq. (3)), which is an estimator of the LWS of an image. In fact, as mentioned in Sec. 2.1, the LWP neglects the constant component of the image contained in the father wavelet coefficients. Based on these results, we avoid adding a stochastic structure to deterministic images in the following analysis.

The stationarity analysis has also been repeated on modified versions of the chessboard image in order to understand how various types of distortions affect $\mathbf{p}$. First, the image is stretched in the horizontal and vertical directions, see Figs. 1b-1i. Their p-values appear to be always $= 1$ and, thus, unaffected by the stretching. The corresponding $\mathbf{p}$ is also mainly unitary. However, note the repeatable effect on scales 2^1 to 2^2 shown in Fig. 2, obtained from Figs. 1c and 1g. This effect arises from $T_{ave}\{\hat{\mathbf{S}}\}$ having a peak in the diagonal direction which is stronger than in any other image. This peak is present in every image, at scale 2^2 in the horizontal and vertical directions and 2^3 in the diagonal one. However, in the case of Figs. 1c and 1g, it is particularly strong in respect with the values at other scales. In fact, in these levels and with these images, the trade-off between spatial distance and frequency of the changes in intensity is the highest. Given that the stretched images appear visually stationary, this unexpected effect highlights a limitation of the mathematical method.

Next, we rotated the chessboard image to confirm the independence of the test from the direction of the texture. This should be ensured by the completeness of the family of filters $\{\psi_j^l\}$ considered, as they probe all the relationships between pixels at the scale j. We used the function `imutils.rotate` from Python,[22] based on bilinear interpolation, with rotation angles $[0°, 90°)$, in an interval of $10°$. Bigger angles of rotation are not necessary, given that the original chessboard image is symmetric by rotation of π and the horizontal and vertical wavelets probe perpendicular directions. As expected, all these images appeared to be data stationary at every scale considered for each rotation angle.

Finally, we varied the sizes of the white and black patches in the chessboard. To maintain the original image dimension (128×128), the sides of the patches were enlarged by powers of 2, as shown in Fig. 3. Again, all the images appear to be stationary with $\alpha = 5\%$. However, it is interesting to notice that the p is not unitary for all the images, as shown in Fig. 4. Visually, this

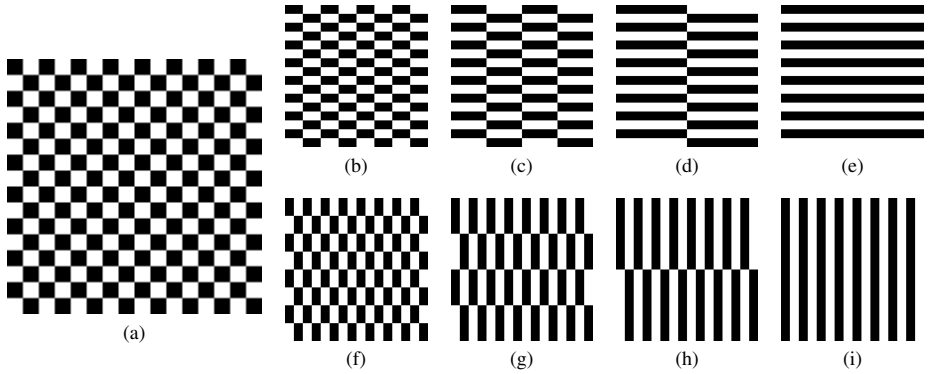


Fig 1: Original chessboard image (Fig. 1a) and horizontally (Figs. 1b, 1c, 1d and 1e) and vertically (Figs. 1f, 1g, 1h and 1i) stretched versions.

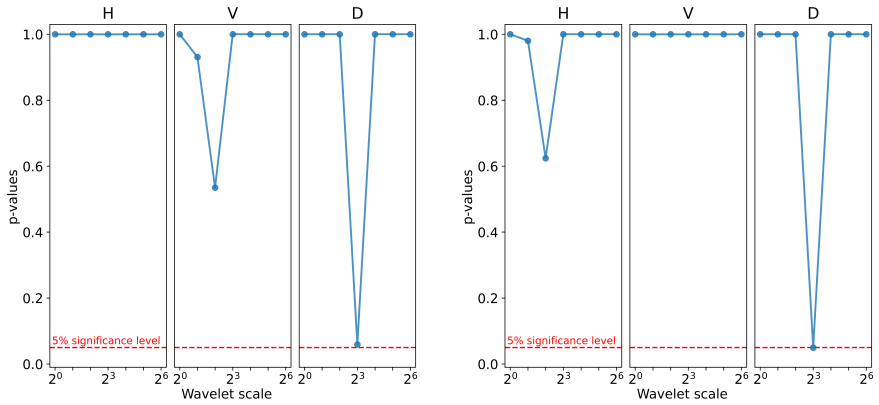


Fig 2: ps for Figs. 1c (left) and 1g (right) (H for horizontal, V for vertical, and D for diagonal).

can be linked to a reduction of homogeneity of the whole image associated with the scaling. In parallel, an analysis of the vector $\mathbf{p}$ of the images, shown in Fig. 5, displays an average decrease of values with the zoom, especially in the finer scales. This could be due to the different Cone Of Influence (COI) of the wavelet at each scale: the finer wavelets, whose COIs are the smallest considered, are affected by the edges of the pattern more abruptly than the other scales, which could lead to a smaller stationarity.

4 Texture Classification

To further probe how the stationarity information can be linked to the perception of texture we ran a classification experiment. We used the Describable Textures Dataset (DTD),[10,23] whose purpose is

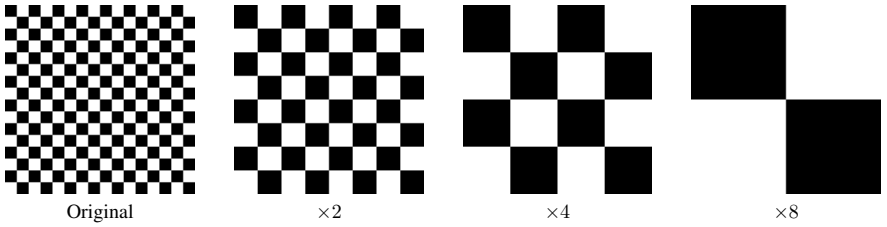


Fig 3: Original image and resized variants.

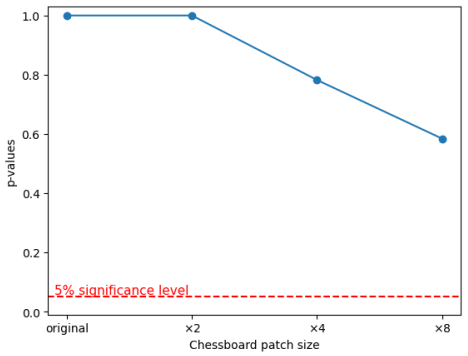


Fig 4: Dependence of chessboard image p-value on the size of its patches.

to describe textures 'in the wild' with semantic attributes chosen by human observers. We selected 100 texture images among the ones in the database, extracting them from 10 different classes derived from Ref. 5. These have been then supplemented with an eleventh class, consisting of 10 pictures of fabric samples which we acquired ourselves, for a total of 110 images. The limited number of images considered is bounded by the necessity of submitting them to the observers of the psychophysical experiment discussed in Sec. 5. The classes are: *chequered*, *dotted*, *fabric*, *flecked*, *grid*, *knitted*, *lacelike*, *scaly*, *stratified*, *striped* and *waffled*. While the original images are in colour, in this experiment they have been converted to greyscale images, in order to account only for their spatial variation. This conversion has been performed by loading each image with the `cv2.imread` function of the OpenCV library, which derives the intensity information Y as $Y = 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B$.[24] Every picture has also been cropped into squares of 128×128 pixels, as required by the current implementation of the algorithm. Examples of the selected and processed images from all 11 classes are shown in Fig. 6.

We ran two classification tasks on these data. For the first task, we divided each image input into sub-images. To comply with the requirements of the dyadic implementation of LS2W, each image target was split into 16 sub-images, each of dimension 32×32 pixels (for additional insight on the choice of dimension, see Appendix D). Then, we classify these sub-images and evaluate whether they are classified as belonging to their original image. Details on the experiment setup for the first task can be seen in Fig. 7. As for the second task, all the 110 images were classified

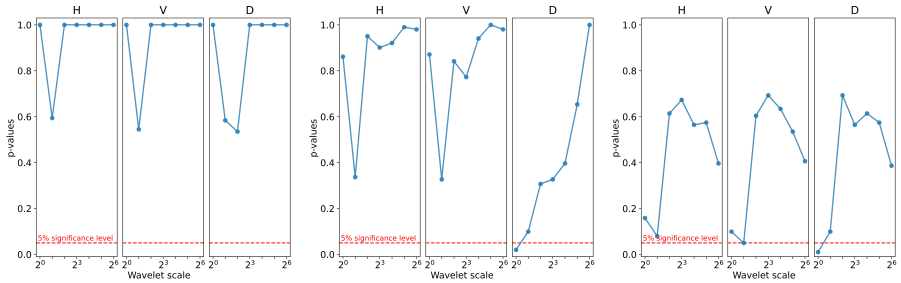


Fig 5: Vectors $\mathbf{p}$ for $\times 2$, $\times 4$ and $\times 8$ resized chessboards showed in Fig. 3. Note that the y-axes are identical and the labels over each graph indicate the direction of the wavelet (H for horizontal, V for vertical, and D for diagonal).

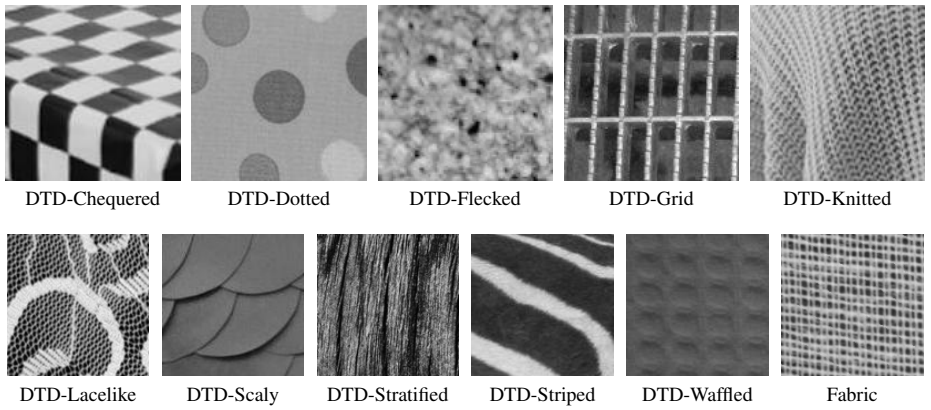


Fig 6: Examples of images used in the experiment, coming from all 11 classes or categories of images. 10 classes originate from DTD[10] and one comes from our own dataset of white fabrics.

using more varied texture features and the previously mentioned DTD groups as ground truth classes. Details for this experiment are given in Sec. 4.3.

4.1 Texture Feature Extraction

Before presenting the results of the classification tasks, we introduce the texture features used for them. Five texture feature vectors are extracted from each image, each considered at the seven dyadic scales (five for the sub-images) used to extract its corresponding $\mathbf{p}$. Feature extraction methods have been selected from the collection of techniques considered in Ref. 6. We neglected the non-scalable approaches and the learning-based ones. We did not consider the latter because, without additional training, the off-the-shelf learning models are also non-scalable, and also because they have not been discussed by Ref. 3 when defining stationary textures. However, we tested

as reference the capabilities of off-the-shelf models in Sec. 4.3. Details of the feature extraction approaches used and their parameters are as follows.

Rotation-invariant local binary pattern (LBP) [25] vectors are obtained at dyadic radii of $2^n,\ n \in [0, 7]$ and eight angles $(k \cdot \pi/4)$ with interpolation. For each scale and image, an LBP histogram of 36 elements is then used as its feature vector.

Grey-level co-occurrences matrix (GLCM) [26] vectors are obtained. They are composed of five global statistical parameters (contrast, correlation, energy, entropy and homogeneity) at four angles (0, $\pi/4$, $\pi/2$, $3/4\pi$). The feature vector is therefore 20 elements long for each scale.

Histogram of oriented gradients (HOG) [27] is computed at nine orientations, normalized according to the hysteresis L2-norm, with 2^n pixels per cell and one cell per block. The HOG vectors thus obtained have a total of nine elements per scale.

Gabor filters-based features [28] with central filter frequencies $f = 1/2^n,\ n \in [0, 7]$, eight orientations, and deviation parameters γ and η assumed to be equal $3\ln(2)/2\pi$. γ and η are chosen such that half-peak magnitude iso-ellipses of the various filters would not overlap (see Appendix C of Ref. 29 for more details). Input images are filtered with these and their mean and variance has been calculated, resulting in a 16-element vector for each scale.

Wavelet vectors, generated with Haar and Daubechies filters at dyadic scales. These features correspond to methods I and III used in Ref. 17. As in the Gabor-based ones above, the mean and variance energy of the filtered images have been used. Note that the variance of the energy has some degree of correlation with the LWP (Eq. (3)) and with the test statistics adopted (Eq. (4) and Eq. (5)). However, the two have some substantial differences: while the regular wavelets used to extract the features are placed at dyadic locations in the image, the LS2W model used by the test is based on non-decimated discrete wavelets. Moreover, the estimator $\hat{\mathbf{S}}(\mathbf{u})$ is corrected with the discrete autocorrelation wavelet matrix A_J. Horizontal, vertical, and diagonal wavelets have been considered, such that each feature vector has a length of 12.

According to Ref. 3 and 6, within this list, GLCM and HOG are texture features better suited for stationary texture images, while the others are better at characterizing non-stationary ones. This suggests that it could be possible to evaluate the stationarity of an image based on the other features. This however could be achieved only by defining a proper testing procedure for each methodology.

4.2 Classification Task 1: Mixed Sub-image Classification

The vector $\mathbf{p}$ by itself reflects the stationarity (or lack of it) of an image. Therefore it is not able to wholly represent the peculiar characteristics of a texture by itself, which is what is required by the features used for classification. Nonetheless, it is possible to use the stationarity information to optimize the process of texture feature extraction. In fact, some feature extraction techniques are claimed to be more appropriate for non-stationary images than others,[3,6] although such a claim has not been proven experimentally. According to this idea, the stationarity information contained in the $\mathbf{p}$ could suggest a selection of features at different scales which is optimal to describe the

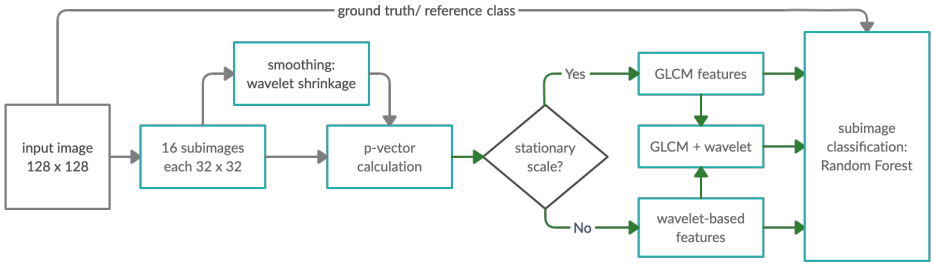


Fig 7: Experiment setup for the mixed sub-image classification task in Sec. 4.2. In it, each input image is split into sub-images, and the latter is classified and evaluated as whether they belong to their original or source image.

texture. In the context of classification, this translates into an increase in the accuracy of the process. In this section we set out to provide experimental proof for this hypothesis.

The experiment setup for this classification task can be seen in Fig. 7. The choice of features for stationary and non-stationary texture is based on Ref. 3. Both GLCM and wavelet-based approaches were used to extract features at dyadic scales, allowing us to classify the sub-images using each individual feature vector. We also combine GLCM and wavelet-based features based on the $\mathbf{p}$ of each image. One of the methods, indicated with f_s, is applied to stationary scales and the other, referred to with f_{ns}, to non-stationary ones. The j^{th} element of the mixed feature vector f_{mix} is then obtained as:

$$f_{mix,j}(f_s, f_{ns}) = \begin{cases} pad(f_{s,j}) & \text{if } p_j > \alpha \\ pad(f_{ns,j}) & \text{if } p_j < \alpha \end{cases} \tag{6}$$

, where $f_s \neq f_{ns}$ and $p_j \in \mathbf{p}_j$. The hypothesis behind the calculation of f_{mix} is that the only features affected by the non-stationarity of a certain scale would be the ones at that same scale. In the present case $f_s = GLCM$ and $f_{ns} = wavelet$. The threshold α, which in the current work is set to 5%, is applied to the values p_j to estimate whether each scale is stationary. If the p-value at a certain scale is bigger than α, the image is considered stationary at that scale and the GLCM feature vector is inserted in the mixed vector. Otherwise, the image at that scale is considered non-stationary and the wavelet-based feature is used. In this way, the space of the mixed vector can be divided into a stationary sub-space and a non-stationary one, each one orthogonal to the other. As a note, $f_{mix,j}$ is padded with zeros to the right in the stationary case and to the left in the non-stationary one, so that the length of the mixed vector is equal to the sum of the lengths of the other two vectors.

As shown in Fig. 7, we chose to use a Random Forest classifier with a $67\% - 33\%$ training-test set subdivision. The forest has 100 trees, and the algorithm selects their depth so that the nodes are expanded until all the leaves are pure. At each split of the tree, the square root of the initial number of features is considered. For every process, the classification has been repeated 1000 times and the average value of the accuracy was extracted. The results of using GLCM and wavelet-based features individually as well as in a mixed feature vector are shown in Table 2. Combining the two techniques appears to worsen the classification accuracy. The use of wavelet as f_s and GLCM

as f_{ns} is discussed in Appendix B. Appendix B also shows how the use of wavelet shrinkage in the extraction of the $\mathbf{p}$, discussed in Sec. 2.1, leads the accuracy to drop to 61.81%. Additional analyses of alternative classification experiments are also reported in Appendices A and E.

Table 2: Classification accuracy corresponding to the classification task 1 shown in Fig. 7.

Method	GLCM	Wavelet	f_{mix}
Average accuracy (in %)	68.05	65.33	64.62

4.3 Classification Task 2: Mixed DTD Classification

As a second task, we classified the unabridged images on the basis of the classes defined by the DTD authors.[5] In this case, we used all the features extraction techniques described in Sec. 4.1, so to probe a wider range of possible approaches. As in the previous section, we derived the accuracy obtained both by classifying the dataset with the original features and with all the possible combinations of mixed vectors f_{mix} (see Eq. (6)). The results for the classification without shrinkage are reported in Table 3, where, according to the practice adopted in Eq. (6), f_s indicates the feature extraction method considered as stationary, while f_{ns} the non-stationary one. On the diagonal of the tables we show as reference the results for the classification without mixing. The results of applying shrinkage to the same set of experiments are also shown in Table 4.

These results are compatible with those obtained by local descriptors for the whole DTD.[10] Between the unmixed features of Tables 3 and 4, the Local Binary Patterns are the most successful. On the other hand, HOG and Gabor features appear to perform quite poorly. The mix that provides the best classification accuracy is $f_s = wavelet$ and $f_{ns} = LBP$. In general, the mixing appears to improve the performance of the classification with every technique. In this case, applying the wavelet shrinkage when calculating $\mathbf{p}_j$ seems to be the best choice.

The accuracy depends more on the stationary technique (f_s) than on the non-stationary one (f_{ns}). This is due to the fact that the images chosen are mainly stationary: the 88% of the p_js are bigger than the 5% test threshold without wavelet shrinkage, while the percentage drops to 77% with wavelet shrinkage. The whole-image p-values have a similar statistic, with 90% for the rough and 92% for the smooth. This could be related to the fact that the test used is conservative.[8]

As an additional reference, we performed the same experiment with the following seven Convolutional Neural Networks (CNNs):

- ResNet-50[30]
- VGG-16 and VGG-19[31]
- Inception v3[32]
- DenseNet-121, DenseNet-161 and DenseNet-201[33]

We extracted the features from off-the-shelf models, which were trained for object recognition. Each network was employed as a generic feature extractor, and the resulting features were then passed on to the Random Forest classifier. Every individual network extracts 1000 features per image. Given that the CNNs required the input array to have certain dimensions and to have three channels, we resized them accordingly using cubic interpolation and we tripled the gray channel.

Table 3: Classification accuracy corresponding to the classification task 2 without wavelet shrinkage.

		f_{ns}				
		LBP	GLCM	HOG	Gabor	wavelet
f_s	LBP	32.7%	32.4%	31.6%	32.8%	32.3%
	GLCM	32.5%	31.7%	31.3%	31.4%	31.9%
	HOG	27.5%	26.2%	23.4%	25.9%	25.9%
	Gabor	27.4%	27.3%	27.6%	26.2%	27.3%
	wavelet	33.8%	32.0%	31.4%	32.5%	29.9%

Table 4: Classification accuracy corresponding to the classification task 2 with wavelet shrinkage.

		f_{ns}				
		LBP	GLCM	HOG	Gabor	wavelet
f_s	LBP	32.7%	34.0%	30.3%	30.2%	33.6%
	GLCM	33.8%	31.7%	31.6%	29.2%	33.3%
	HOG	33.7%	33.2%	23.4%	26.4%	33.1%
	Gabor	34.7%	32.7%	33.9%	26.2%	32.4%
	wavelet	36.6%	34.4%	31.1%	33.9%	29.9%

The results are shown in Tab. 5, where one can see that this assignment is challenging even for learning-based techniques.

Table 5: Classification accuracy with CNNs corresponding to the classification task 2.

Network	Average accuracy (in %)
ResNet-50	44.0
VGG-16	39.7
VGG-19	47.4
Inception v3	40.5
DenseNet-121	43.3
DenseNet-169	42.0
DenseNet-201	42.7

5 Psychovisual Experiment Design

As final step of our investigation of the link between perceptual and data stationarity, we performed a psychophysical experiment. We used the psychovisual software PsychoPy2[34] to set up the experiment and uploaded it to the Pavlovia web platform. The images used in the experiment are the same 110 which were classified in the previous section, greyscaled and cropped. The experiment was performed by 93 observers, who carried it out on their personal computer and screen. Therefore, the viewing environment of each observer was uncontrolled, which could pose some challenges, mainly related to the resolution of the image, which will vary with the type of display and the distance of the observer from the screen. However, this effect is limited by the fact that Pavlovia automatically activates the full-screen view when the experiment starts. As the images

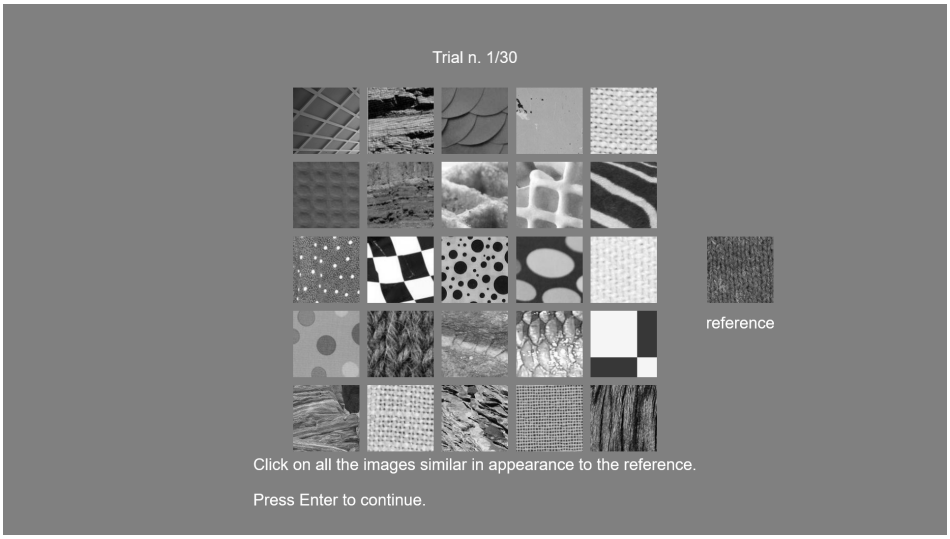


Fig 8: A screenshot of the psychophysical experiment performed.

were greyscale, a colour calibration of the screen was not necessary. The display settings of each observer could have had an impact, although as discussed in Ref. 35, many studies have compared online behavioral experiments with lab-based ones, and they found that their data quality is usually equivalent.

The experiment was divided into 30 rounds. At each of them, an observer was presented a texture reference and 25 samples, and was asked to select all the images that looked similar to the reference. No information other than this was provided to the observers, and no definition of the words *similar* and *texture* was given before the experiment. The 26 textures, samples and reference, were selected randomly from the database, and therefore rounds without instances of the reference image class in the 25 samples were possible. An example of an experiment round is shown in Fig. 8. Based on the results of this experiment, it is possible to evaluate how similar two texture images A and B are by defining a similarity coefficient $SIM_{A,B}$ (Eq. (7)). Here, $n_{group}(A,B)$ indicates the number of times that A and B have been grouped together, while $n_{appear}(A,B)$ the number of times they appeared together in the same screen, given that A is a reference image. Note that $SIM_{A,B} \in [0,1]$.

$$SIM_{A,B} = \frac{n_{group}(A,B)}{n_{appear}(A,B)} \tag{7}$$

The results of this similarity evaluation process can be used to fill a matrix, as shown in Fig. 9. In this figure, we highlighted the boundaries between images belonging to different DTD classes. From the figure, it can be seen which classes are confused with each other, such as the *chequered* with the *grid*, the *flecked* with the *dotted* and the *knitted*, and the *scaly* with the *stratified*.

If averaged over each class, this similarity coefficient matrix can be compared to the correlation matrix of the classification task 2, in Sec. 4.3. This is justified by the observation that, if

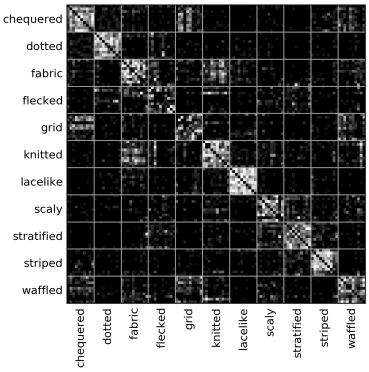


Fig 9: The $SIM_{A,B}$ matrix for the experiment performed.

Table 6: Spearman's rank correlation coefficients between confusion matrix of the classification with different feature extraction methods and the $SIM_{A,B}$ matrix (Fig. 9).

Method	LBP	GLCM	HOG	Gabor	Wavelet
Spearman's ρ	0.51	0.49	0.47	0.48	0.38

the images belonging to a certain class are generally similar to those belonging to another according to the average human observer, it is more likely that the classification algorithm will confuse them. Therefore, we calculated the average confusion matrix of 1000 classification repetitions for the feature extraction techniques used in Sec. 4.3, and we then calculated the Spearman's rank correlation coefficient ρ between the two matrices after having collapsed them to one-dimensional vectors. We chose to use this measure because we want to evaluate the relationship between the two elements, without assumptions about its linearity or the type of distributions the data is obtained from. The results are shown in Table 6. The values of the ρs show a moderate correlation and they seem to reflect the accuracies obtained in the second classification task. The only exception to this rule are the wavelet features, which show relatively low ρ, but give a relatively high classification accuracy in the group. The same analysis can be performed with the mixed features. Without wavelet shrinkage we obtain the results shown in Table 7. We have also conducted the experiment with wavelet shrinkage, resulting in ρs on average 0.08 smaller than the ones without it. Interestingly, this shows how, while the best solution to classify images seems to be a mix of wavelet and LBP features vector, the results that better fit with the human observation are obtained by mixing wavelets with GLCM.

Another way to link this psychovisual similarity to the inspected features is by comparing the difference between the feature vectors of each pair of images with their similarity. This, too, can be gauged with the Spearman's ρ. First, we extracted the feature vectors as described in Sec. 4.1, calculated the distance between each of them, and compared the output with the similarity values. The results are provided in Table 8. Notice that they are all negative and quite small.

If we do the same for the mixed features, without wavelet shrinkage we get results shown in Table 9. If we add wavelet shrinkage, the correlation coefficients are on average the same. The best possible choice of feature mixing is, in this case, the pure co-occurrence matrix features. On the

Table 7: Spearman's rank correlation coefficients between confusion matrix of the classification with different stationarity-based mixed methods and the $SIM_{A,B}$ matrix (Fig. 9).

		Non-stationary				
		LBP	GLCM	HOG	Gabor	wavelet
Stationary	LBP	0.51	0.47	0.45	0.46	0.49
	GLCM	0.53	0.49	0.52	0.51	0.53
	HOG	0.46	0.46	0.47	0.43	0.47
	Gabor	0.49	0.49	0.5	0.48	0.48
	wavelet	0.38	0.34	0.34	0.35	0.38

Table 8: Spearman's rank correlation coefficients between feature vector distances and the $SIM_{A,B}$ matrix (Fig. 9).

Method	LBP	GLCM	HOG	Gabor	Wavelet
Spearman's ρ	-0.15	-0.18	-0.1	-0.15	-0.11

other hand, wavelets appear to perform extremely poorly as stationary features. However, overall, the Spearman's ρ indicates that the correlation between the techniques used and the results of the psychovisual experiment performed is very weak. This is in line with the results of Sec. 4.3, which shows how demanding the DTD classification's task is.

Table 9: Spearman's rank correlation coefficients between distances of feature vectors obtained with different stationarity-based mixed methods and the $SIM_{A,B}$ matrix (Fig. 9).

		Non-stationary				
		LBP	GLCM	HOG	Gabor	wavelet
Stationary	LBP	-0.15	-0.12	-0.15	-0.12	-0.13
	GLCM	-0.11	-0.18	-0.11	-0.15	-0.15
	HOG	-0.13	-0.11	-0.1	-0.11	-0.12
	Gabor	-0.16	-0.17	-0.16	-0.15	-0.13
	wavelet	-0.11	-0.11	-0.11	-0.11	-0.11

6 Discussion and Conclusion

The results obtained in this work provide clues on how data stationarity is linked to human perception. First of all, the analysis of the chessboard images in Sec. 3 demonstrated its fundamental properties in relation to the simple case of a regular texture. According to the test used, the original chessboard image is data stationary, as are its stretched variations (Fig. 2). On the other hand, an increase of the chessboard patches dimension reduces the stationarity, particularly at lower scales and higher spatial frequencies (Fig. 5).

The classification experiments, described in Sec. 4, provide us with additional insights in relation to irregular textures. In the first task, discussed in Sec. 4.2, the p_js, i.e. the p-values resulting from testing a texture for stationarity at different scales, are used to mix the elements of GLCM and wavelet vectors. This however does not improve the classification accuracy. Additional analyses are reported in the Appendices. The results of this classification task suggest that Petrou and

Sevilla's claim that while model-based texture features like GLCM are more suited for stationary images, frequency-based descriptors such as wavelet are preferred for non-stationary ones[3] has a limited validity in a classification framework. On the other hand, the second task, reported in Sec. 4.3, shows that using the stationarity information does improve the classification of the DTD images in their texture macrogroups. Compared to task 1, this assignment is more related to high level texture perception.

Finally, the psychophysical experiment (Sec. 5) directly probes the link between perception and math. In its first part, it addresses the correlation between the confusion matrices of Sec. 4.3 and the similarity results, revealing how a mix of wavelet and LBP best replicates the average observer's response. It also demonstrates that the traditional texture features are very weakly correlated with the results of the visual experiment. Even in this context, despite the small size of the Spearman's ρs, the $\mathbf{p}$-based mixing of features increases it.

In general, it is not clear if using wavelet shrinkage during the $\mathbf{p}$ calculation improves or reduces the relationship between visual and data stationarity. In some cases applying the wavelet shrinkage to $\mathbf{p}$ has a disruptive effect, while in some others the effect is negligible. For example, in the chessboard experiment of Sec. 3 the shrinkage filters high-frequency artifacts, which degrade some $p_{\eta(j,l)}$s of the $\mathbf{p}$, and it is therefore convenient. In the experiment of Appendix B its application reduces the accuracy of the algorithm, while it increases it in Appendix E. Finally, it has a negative effect on the psychovisual analysis between mixed features and experimental similarity.

To conclude, the current paper shows how stationarity information can be linked to the psychophysical attributes of a texture image and how evaluating it with Locally Stationary Two-Dimensional Wavelet (LS2W) processes can be used to improve a texture classification pipeline. There are various possible future steps that could better clarify the role of stationarity in texture. Firstly, one can examine a wider variety of spatial stationarity metrics.[36–38] Even if the disadvantages of most of these have been highlighted in Ref. 8, their relationship with the perception of texture has yet to be assessed. The LS2W method itself can be improved. The scale analysis is currently performed at dyadic scales,[20] which allows fast extraction of the p of an image, but obtaining $\mathbf{p}$ on a continuous range of scales would provide more insight on its behaviour. Another possible improvement to this approach is to expand it to colour and spectral images and to find the best way to mix the various image channels. It has, in fact, been demonstrated that taking them into account increases the performance of texture analysis.[39] This has been already done for p in Ref. 21, but not for scale-dependent $\mathbf{p}$. Finally, p, $\mathbf{p}$ and $\mathbf{p}_j$ can be used to detect which image in a database has to go through a texture segmentation process.

Appendix A: Sub-image classification using GLCM and wavelet-based features at all scales

In Sec. 4.2, we calculated the accuracy obtained by classifying images with f_{mix}, a feature vector obtained by mixing GLCM and wavelet elements. In this appendix, we compare the results of that experiment, shown in Table 2, to what can be achieved by simply combining the GLCM and wavelet vectors as in Eq. (8). Here, c is an operation concatenating the vector $f_{ns,j}$ to $f_{s,j}$. With this, we achieve an accuracy of 72.3%, which is the best accuracy reached for this experiment (see Table 2).

$$f_{comb,j}(f_s, f_{ns}) = c(f_{s,j}, f_{ns,j}) \tag{8}$$

Appendix B: Sub-image classification using variations of f_{mix}

For the classification based on f_{mix} (see Sec. 4.2) we also considered the cases in which $\mathbf{p}$ is calculated without wavelet shrinkage-based smoothing. Ref. 17 demonstrates that the results of the stationarity test are more reproducible if smoothing is applied, but this could be counterproductive for the classification. Moreover, we assessed the case in which $f_s = wavelet$ and $f_{ns} = GLCM$ (see Eq. (6)), since this would provide us with experimental proof for the considerations proposed in Ref. 3, i.e. that the some techniques, like GLCM, are more suited to stationary images than others, such as wavelets.

Table 10: Classification accuracy corresponding to the classification task 1 shown in Fig. 7, with various choices of stationary features and with wavelet shrinkage.

Shrinkage	Stationary	Non-stationary	Accuracy
no	GLCM	wavelet	64.62%
no	wavelet	GLCM	61.18%
yes	GLCM	wavelet	61.81%
yes	wavelet	GLCM	57.04%

The results obtained by calculating the p_js of each sub-image are shown in Table 10. We can see that the 'rough' $\mathbf{p}_j$ that is obtained without applying wavelet shrinkage perform better. These outputs are partially in line with Ref. 3's hypothesis discussed in Sec. 4.1, as using wavelet as stationary technique is worse than using GLCM. However, the accuracies reported in Table 10 are all smaller than those obtained with pure GLCM and wavelet features (see Table 2) and with a combination of the two (see Appendix A).

Appendix C: ALOT analysis

As discussed in Sec. 1, we chose to use the DTD images in our analysis because of their vision-based arrangement. To provide an alternative, we also considered the Amsterdam Library of Textures (ALOT).[40] In particular, we studied the ALOT pictures mentioned in Ref. 6. In this article, the authors classified a wide variety of existing texture databases according to certain characteristics, among which is texture stationarity, with explicit mention to the definition given in Ref. 3. Therefore, we applied the LS2W stationarity test (Sec. 2) to two datasets defined in this paper and extracted from the ALOT: one stationary `ALOT-95-S-N` and the non-stationary `ALOT-40-NS-N`. Setting the significance level α to 0.1 and the number of bootstrap iterations B to 10, only 40% of the `ALOT-95-S-N` images are classified as stationary by the test, whereas for the `ALOT-40-NS-N` group this percentage is increased to 67.5%. This demonstrates the need for a common definition of data and visual stationarity.

Subsequently, we expanded the results reported in Sec. 4.2 by applying the first classification task to the ALOT. In particular, we merged the two classes `ALOT-95-S-N` and `ALOT-40-NS-N`. We adopted the same approach as Sec. 4.2, dividing each sample in 16 sub-images of shape 256×256. The results of the classification are shown in Table 11, where $f_{mix,ws}$ indicates the mixed features obtained without wavelet shrinkage and $f_{mix,s}$ the ones with it. These results are similar to those attained in Sec. 4.2.

Table 11: Classification accuracy of task 1 performed on the ALOT images.

Method	GLCM	Wavelet	$f_{mix,ws}$	$f_{mix,s}$
Average accuracy without shrinking (in %)	72.18	65.99	70.46	71.92

Appendix D: Dimension dependence

In Sec. 4 we divided each image in sub-images of dimension 32×32. This is in contrast with the results discussed in Ref. 8, whose experiment on the power assessment of the LS2W test shows that an image size of at least 128 pixels side is required to achieve good statistical power. However, this conclusion has been obtained based on artificial non-stationary models whose visual non-stationarity is extremely low (e.g. see Fig. $3d$ in the reference). Moreover, the chosen sub-images' size is limited by the dimension of the images selected from the DTD.

In this Appendix, we analyse how the size of the sub-images can influence the results of Sec. 4.2. First, we divided the DTD images selected into bigger sub-images. This has the drawback of reducing the total number of images available for the classification. We then repeated task 1 of the classification section. With sub-images of size 64×64, which correspond to dividing the original picture into four squared sections, we obtain Table 12, while with sub-images of size 42×42, we obtained Table 13.

Table 12: Classification accuracy of task 1 performed on images with size 64×64.

Method	GLCM	Wavelet	f_{mix}
Average accuracy (in %)	63.55	58.84	57.62

Table 13: Classification accuracy of task 1 performed on images with size 42×42.

Method	GLCM	Wavelet	f_{mix}
Average accuracy (in %)	68.77	67.15	62.32

As in Sec. 4.2, we reported the numbers obtained by mixing the features using the p_js obtained without wavelet shrinkage, as applying it would slightly reduce the classification performance. One can see that the results obtained in Sec. 4.2 correspond, with minor variations, to the ones showed here.

As mentioned, Ref. 8 suggests using square images with sides of at least 128 pixels. To satisfy this requirement without reducing the number of samples for the classification, we randomly extracted sub-pictures of size 128×128 from the selected DTD images. For each image, we derived 16 sub-pictures, so that the number would correspond to the batch used in the calculation of Table 2, for a total of 1760 samples. The output of this experiment is shown in Table 14. In this case the accuracy is strongly enhanced, probably due to the fact that it is likely that some of the classified pictures overlap. Nonetheless, the conclusions of Sec. 4.2 are still unaffected by the change of dimension of the images. In this case, mixing with wavelet shrinkage is the best performing method, and thus it is the number reported in Table 14.

Appendix E: Sub-image classification using image source $\mathbf{p}$

In Sec. 4.2, we extracted the mixed features vector f_{mix} using a different set of p_js for each sub-image. If we repeat the experiment with a common $\mathbf{p}_j$ for all sub-images belonging to the same

Table 14: Classification accuracy of task 1 performed on images with size 128×128.

Method	GLCM	Wavelet	f_{mix}
Average accuracy (in %)	95.26	95.01	93.19

original image, we get the results shown in Table 15. Here, we can see a clear improvement with respect to the case discussed in Appendix B, due to the fact that sub-images with common origin have the same null terms. Nonetheless, the improvement is quite significant and it is interesting how the wavelet shrinkage further boosts it. In this case, the classification accuracy is actually improved in respect with the output obtained using pure features.

Table 15: Classification accuracy corresponding to the classification task 2 shown in Fig. 7, with common set of p_js.

Shrinkage	Stationary	Non-stationary	Accuracy
no	GLCM	wavelet	77.79%
no	wavelet	GLCM	75.34%
yes	GLCM	wavelet	79.96%
yes	wavelet	GLCM	75.24%

Acknowledgments

This work is partially supported by FRIPRO FRINATEK Metrological texture analysis for hyperspectral images (projectnr. 274881) funded by the Research Council of Norway.

References

1 I. Florescu, *Probability and Stochastic Processes*, John Wiley & Sons, Inc., Hoboken, New Jersey. (2014).

2 H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Transactions on Systems, Man and Cybernetics* **8**(6), 460–473 (1978).

3 M. Petrou and P. G. Sevilla, *Image Processing: Dealing with Texture*, John Wiley & Sons, Inc., Chichester, England (2006).

4 M. Petrou and S. ichiro Kamata, *Image Processing: Dealing with Texture*, John Wiley & Sons, Inc., Chichester, England (2021).

5 N. Bhushan, A. R. Rao, and G. L. Lohse, "The texture lexicon: Understanding the categorization of visual texture terms and their relationship to texture images," *Cognitive Science* **21**, 219–246 (1997).

6 R. Bello-Cerezo, F. Bianconi, F. D. Maria, *et al.*, "Comparative evaluation of hand-crafted image descriptors vs. off-the-shelf CNN-based features for colour texture classification under ideal and realistic conditions," *Applied Sciences* **9**, 738 (2019).

7 M. Conni and H. Deborah, "Texture stationarity evaluation with local wavelet spectrum," in *London Imaging Meeting*, **2020**, 24–27 (2020).

8 S. L. Taylor, I. A. Eckley, and M. A. Nunes, "A test of stationarity for textured images," *Technometrics* **56**, 291–301 (2014).

9 E. V. Kurmyshev, M. Poterasu, and J. T. Guillen-Bonilla, "Image scale determination for optimal texture classification using coordinated clusters representation," *Applied Optics* **46**, 1467–1476 (2007).

10 M. Cimpoi, S. Maji, I. Kokkinos, *et al.*, "Describing textures in the wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 3606–3613 (2014).

11 B. Julesz, "Experiments in the visual perception of texture," *Scientific American* **232**, 34–43 (1975).

12 E. V. Kurmyshev and M. A. Cervantes, "A quasi-statistical approach to digital binary image representation," *Revista Mexicana de Fisica* **42**, 104–116 (1996).

13 B. Vidakovic, *Statistical Modeling by Wavelets*, John Wiley & Sons, New York (2009).

14 C. K. Chui, *An Introduction to Wavelets*, Academic Press, San Diego, CA (1992).

15 S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA (1999).

16 G. P. Nason, R. V. Sachs, and G. Kroisandt, "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**, 271–292 (2000).

17 I. A. Eckley, G. P. Nason, and R. L. Treloar, "Locally stationary wavelet fields with application to the modelling and analysis of image texture," *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **59**, 595–616 (2010).

18 I. A. Eckley and G. P. Nason, "Efficient computation of the discrete autocorrelation wavelet inner product matrix," *Statistics and Computing* **15**, 83–92 (2005).

19 A. C. Davison and D. V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, Cambridge, UK (1997).

20 M. A. Nunes, S. L. Taylor, and I. A. Eckley, "A multiscale test of spatial stationarity for textured images in R," *The R Journal* **6**, 20–30 (2014).

21 S. L. Taylor, I. A. Eckley, and M. A. Nunes, "Multivariate locally stationary 2D wavelet processes with application to colour texture analysis," *Statistics and Computing* **27**, 1129–1143 (2017).

22 "Source of the imutils Python library." https://github.com/jrosebr1/imutils. accessed November 2020.

23 "Describable Textures Dataset (DTD) website." http://www.robots.ox.ac.uk/~vgg/data/dtd/. accessed September 2020.

24 "Documentation of the `imread` function in the OpenCV library." https://docs.opencv.org/4.5.1/d4/da8/group__imgcodecs.html#imread. accessed March 2021.

25 F. Bianconi, R. Bello-Cerezo, and P. Napoletano, "Improved opponent color local binary patterns: an effective local image descriptor for color texture classification," *Journal of Electronic Imaging* **27**, 1–10 (2017).

26 R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**, 610–621 (1973).

27 N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, **1**, 886–893 (2005).

28 T. Randen and J. H. Husøy, "Filtering for texture classification: a comparative study," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**, 291–310 (1999).
29 F. Bianconi and A. Fernández, "Evaluation of the effects of Gabor filter parameters on texture classification," *Pattern Recognition* **40**, 3325–3335 (2007).
30 K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778 (2016).
31 K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556* (2014).
32 C. Szegedy, V. Vanhoucke, S. Ioffe, *et al.*, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).
33 G. Huang, Z. Liu, L. V. D. Maaten, *et al.*, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708 (2017).
34 J. Peirce, J. R. Gray, S. Simpson, *et al.*, "Psychopy2: Experiments in behavior made easy," *Behavior Research Methods* **51**, 195–203 (2019).
35 M. Sauter, D. Draschkow, and W. Mack, "Building, hosting and recruiting: A brief introduction to running behavioral experiments online," *Brain Sciences* **10**, 251 (2020).
36 A. Ephraty, J. Tabrikian, and H. Messer, "A test for spatial stationarity and applications," in *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, 412–415 (1996).
37 S. Bose and A. Steinhardt, "Invariant tests for spatial stationarity using covariance structure," *IEEE Transactions on Signal Processing* **44**, 1523–1533 (1996).
38 M. Fuentes, "A formal test for nonstationarity of spatial stochastic processes," *Journal of Multivariate Analysis* **96**, 30–54 (2005).
39 E. Cernadas, M. Fernández-Delgado, E. González-Rufino, *et al.*, "Influence of normalization and color space to color texture classification," *Pattern Recognition* **61**, 120–138 (2017).
40 G. J. Burghouts and J.-M. Geusebroek, "Material-specific adaptation of color invariant features," *Pattern Recognition Letters* **30**, 306–313 (2009).

Michele Conni received his BS and MS in engineering physics from the Polytechnic University of Milan (2015), with specialization in Optics and Photonics. He is currently studying for a PhD in computer science at the Norwegian University of Science and Technology (NTNU), in collaboration with Barbieri Electronic, where he works in the research and development group.

Hilda Deborah is a Researcher at NTNU. She received her BSc in Computer Science from University of Indonesia (2010), MSc from Erasmus Mundus Color in Informatics and Media Technology (2013), and PhD in Computer Science from NTNU and University of Poitiers (2016). Her current research interests are hyperspectral imaging and texture analysis.

Peter Nussbaum is an associate professor of colour imaging at the Colour and Visual Computing Laboratory, NTNU, Norway. Dr. Nussbaum received an MSc the Colour & Imaging Institute, University of Derby, GB, in 2002 and completed his PhD degree in imaging science in 2011 from the University of Oslo, Norway.

Phil Green is Professor of Colour Imaging at the Colour and Visual Computing Laboratory, NTNU. He is also Technical Secretary of the International Color Consortium, the body that standardizes the ICC profile format and promotes colour management internationally. Dr. Green received an MSc from the University of Surrey in 1995, and a PhD from the former Colour & Imaging Institute, University of Derby, UK in 2003.

Article E

Michele Conni, Peter Nussbaum, and Phil Green. Textile texture descriptors. *Manuscript under review in a journal*

This article is under review and is therefore not included

Article F

Nadile Nunes de Lima, Michele Conni, Phil Green, and Markus Barbieri. Measurement uncertainty for printed textiles. In *2018 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2018

Measurement uncertainty for printed textiles

Nadile Nunes de Lima
Dept of Computer Science
NTNU
Gjøvik, Norway
Barbieri Electronic
Brixen, Italy
nadile.nlima@gmail.com

Michele Conni
Barbieri Electronic
Brixen, Italy
Dept of Computer Science
NTNU
Gjøvik, Norway
michele.conni@barbierielectronic.com

Phil Green
Dept of Computer Science
NTNU
Gjøvik, Norway
philip.green@ntnu.no

Markus Barbieri
Barbieri electronic
Brixen, Italy
markus.barbieri@barbierielectronic.com

***Abstract*—Measurement repeatability and reproducibility were analysed for four different models of colour measurement instrument on PTFE and on printed and unprinted cotton material. The influence of instrument aperture and sample characteristics on the measurements was analyzed, showing that as expected larger apertures tended to give better repeatability and smaller difference in the inter-model agreement. Measurement repeatability across the different textile samples was in the range of 0.04-0.91 ΔE^*_{ab}, while the effect of different sample positioning was in the range of 0.25-0.88 ΔE^*_{ab}.**

***Index Terms*—colorimetry, textiles, measurement geometry, uncertainty**

I. Introduction

Agreement between colour measurement instruments is facilitated by international standards and guidelines such as ISO 13655 [1], GUM [2] and ASTM E2214 [3], and different industries typically have specific recommended measurement parameters. Standards developed by ISO TC130, including ISO 13655, provide recommendations for printed matter generally, but do not include specific recommendations for textiles, which are traditionally measured with instruments with a 0°:d geometry and a relatively large aperture. Digital printers, of both paper-based and textile substrates, use instrumentation based on a 0°:45°geometry and relatively small aperture.

The textile industry commonly uses instruments with a 0°:d (or d:8°) geometry and larger aperture, according to the recommendations of ISO 105 [4]. In addition, the textile industry, in common with many industries, computes colorimetry from spectral reflectance with a D65 illuminant, while graphic arts (and ICC colour management) have adopted D50.

One advantage of a 0°:d geometry when measuring textured materials is that it is less sensitive to the spatial orientation positioning of the instrument with respect to the sample, which results in more consistent results on anisotropic materials such as textiles. In a 0°:45°(or 45°:0°) instrument, greater variation in measurement is expected on such materials.

In order to contribute to recommendations that can be adopted for process control of printed textiles, this work shows the results of uncertainty analysis of different measurement instruments on digitally printed textiles. The procedures in this study attempted to follow the recommendations of ISO DTS 23031 [5], which although not yet approved and published, provides useful recommendations on assessing the performance of colour measurement instruments.

A. Measurement Uncertainty

According to [6], uncertainty (of measurement) is a parameter associated with the result of a measurement that characterizes the dispersion of the values that could reasonably be attributed to the measurand. It can be divided into two main categories: precision and accuracy [7], [8].

ASTM E2214 [3] was developed to standardize the terminology and procedures used to evaluate color measuring instruments. The terminology used in this article and based on ASTM E2214 are the following [9]–[11]:

- Repeatability: closeness of the agreement between the results of successive measurements of the same substrate carried out under the same conditions of measurement.
- Reproducibility: similar to repeatability, except that some aspect of the measurement conditions has changed.
- Inter-instrument agreement: describes the reproducibility of two or more instruments of identical design.
- Inter-model agreement: expresses the reproducibility of two or more instruments of different design.
- Accuracy: how closely an instrument can conform to the accepted or true value for a given sample.

1) Measurement Uncertainty on Textiles: In the textile industry, color control has always been a prime concern,

2018 Colour and Visual Computing Symposium (CVCS)
978-1-5386-5645-7/18/$31.0 © 2018 IEEE

TABLE I: Overview of the four instruments used in this work

	Instrument 1	**Instrument 2**	**Instrument 3**	**Instrument 4**
Geometry	Circumferential 45°:0° (3-point circumferential illumination)	Circumferential 45°:0° (annular illumination)	0°:45°	d:8°, Diffuse illumination
Geometry standard conformance	ISO 13655:2017, ISO-5-4	ISO 13655:2009	DIN 5033	Unspecified
Aperture	Switchable between 2, 6 and 8 mm	4.5 mm	3 mm	8 mm
Aperture (over or under-filled)	Over-filled	Under-filled	Unspecified	Over-filled
Light source	7 narrow-band LEDs	Gas-filled tungsten	Gas-filled tungsten	Gas-filled tungsten
Detector	Diode array	Diode array	Unspecified	Blue-enhanced silicon photodiodes
Inter-instrument agreement	Avg: 0.5 ΔE^*_{00} Max: 1.0 ΔE^*_{00}	Avg: 0.4 ΔE^*_{94} Max: 1.0 ΔE^*_{94}	0.3 ΔE^*_{ab}	Avg: 0.20 ΔE^*_{ab} Max: 0.40 ΔE^*_{ab}
Spectral range and interval	380nm to 750nm at 10nm	380nm to 730nm at 10nm	400nm to 700nm at 10nm	400nm to 700nm at 10nm
Short-term repeatability	Spot: 0.05 ΔE^*_{00} (standard deviation, 10 measurements made with white BCRA) Scan: $<$0.2 ΔE^*_{00}	0.1 ΔE^*_{94} on white (D50, 2°, mean of 10 measurements every 3 seconds on white)	0.03 ΔE^*_{ab}	0.05 ΔE^*_{ab} on white ceramic (standard deviation)

and tolerances are often very tight. Some of the areas where color control is applied are color fastness, quality control, and characterization of colorants [6]. Material texture has an affect on measurement, so it is of particular interest to explore the effects on measurement reproducibility of the spatial orientation and location of measurement on textile samples.

This study focuses on these repeatability and reproducibility aspects of measurement uncertainty for textiles, and we do not show results for combined or expanded uncertainty with respect to calibrated reference materials. Many different models of colour measurement instruments are used in the textile industry. However, the use of instruments with poor inter-instrument agreement can result in complications in the control of the colour reproduction process [12].

II. Experimental

A. Instruments

Four commercial spectrophotometers (one bench-top and three hand-helds) widely used in the graphic arts industry were analyzed. Table I shows the instruments and their manufacturer specifications. Each instrument was calibrated according to the manufacturers instructions using the supplied white reference prior to each measurement set. Instruments were allowed to warm up and left powered on for the duration of each set. Where both scan and spot measurement modes were available, instruments were operated in spot mode. Instruments 1 to 3 were used in ISO 13655 measurement mode M1 and, for the 0:d instrument the specular component was excluded, which is expected to give a better correlation to the 45°:0°instruments.

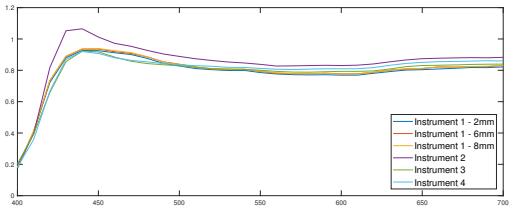
B. Methods and procedures

The performance of the instruments was first evaluated in terms of precision, repeatability, and reproducibility, using a white Spectralon tile, a white paper sample (with optical brightener), and an unprinted sample of each of the fabric samples (see Table II). Textiles have a more inhomogeneous surface than paper substrates due to the threads from which they are woven, and the thread patter varies significantly between different materials.

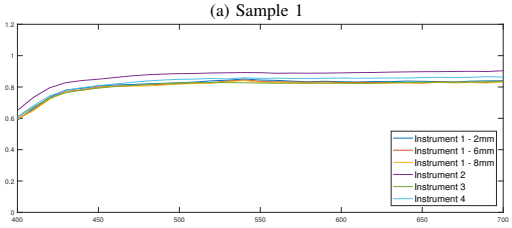
TABLE II: Cotton materials used in the study

	Sample 1	**Sample 2**	**Sample 3**	**Sample 4**	**Sample 5**
Name	Half Panama	Popeline 40/40	Voile	Woven 30/30	Woven 30/22
Thickness	250 microns	237,5 microns	125 microns	150 microns	175 microns
Thread count (per cm^2)	25x20	40x30	36x28	40x30	50x22

Guides such as [13], [14] discuss how to present textile substrate for measurement, but they are focused on yarns and dyed rather than printed textiles. Measurements on paper samples differ from those on real fabrics since they are more or less planar,while fabric surfaces have more varied surface topology [15]. For this study, the fabric samples were mounted on a white backing to provide a stable base substrate for the sample positioning mechanism of the instruments. As preparation for measurements, the instruments were warmed up with random measurements and then calibrated on their own white reference tile supplied by the manufacturer according to the manufacturers instructions. To determine the short-term repeatability, a sequence of 25 readings with a gap of five seconds between each one was taken [1]. The Mean Colour Difference from the Mean (MCDM) [7] for each set of readings has been calculated by equation (1) and the results



(a) Sample 1



(b) Sample 2

Fig. 1: Spectral reflectance of textile samples and the Spectralon reference tile

are shown below, where N is the number of readings, C_i are the colour coordinates of the i^{th} readings, and C_m are the colour coordinates of the average of all readings.

$$MCDM = \frac{1}{N}\sum_{N}\Delta E(c_i, c_m) \tag{1}$$

All instruments report spectral reflectance factor relative to a perfect reflecting diffuser at 10nm intervals. As shown in Table I, the different instruments measure over a different wavelength range, and, the spectral reflectance factor from 400 to 700 nm was used for the computation of colorimetric values, using the procedure described in ISO 13655:2017. Depending on the aperture of the instrument, the number of threads included in the measurement will vary. One instrument in this study has the option of changing the aperture size, and results for all three apertures are reported below for this instrument.

III. Results

A. Short-term repeatability

Fig. 1 shows the spectral reflectance of the textile samples on all four instruments. Sample 1 contains an optical brightening agent and is noticeable the different response to the fluorescent emission between the Instruments 1-4, where Instrument 2 gives a higher prediction of the fluorescent emission. Samples 2-5 have similar spectral data, hence Fig. 1b shows the data for sample 2.

Table III shows manufacturers agreement and the corresponding results in terms of the short-term measurements. To calculate the MCDM, a white pressed PTFE (Spectralon) tile was used. As can be noticed, all instruments had satisfactory repeatability behavior corresponding to the manufacturers tolerances. Fig. 2 shows the MCDM for measurements on the Spectralon tile, the backing paper and the five textile samples. The MCDM values for the textiles are broadly similar for each sample. On sample 1, the instrument 2 shows a discrepancy with respect to the other instruments. In the results for maximum ΔE^*_{ab}, for instrument 1, the 2mm aperture usually has higher values than 6mm and 8mm apertures, particularly on textiles. This is to be expected since a smaller aperture collects light from a smaller area of the sample and is more prone to measurement noise.

TABLE III: Short-term repeatability

	Short term repeatability (manufacturer's agreement)	**Short term repeatability (MCDM measured)**
Inst.1	Spot: 0.05 ΔE^*_{00} (standard deviation, 10 measurements made with white BCRA)	0.05
Inst. 2	0.1 ΔE^*_{94} on white (D50, 2°, mean of 10 measurements every 3 seconds on white)	0.05
Inst. 3	0.03 ΔE^*_{ab}	0.03
Inst. 4	0.05 ΔE^*_{ab} on white ceramic (standard deviation)	0.03

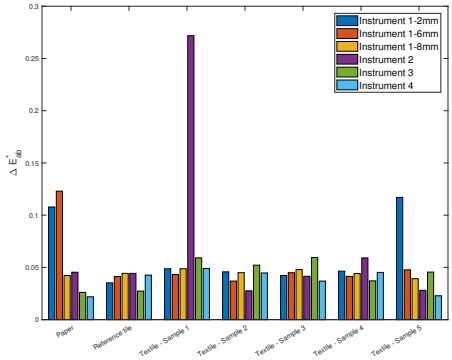


Fig. 2: MCDM for each set of measurements

All five textile samples are 100% cotton, but with different structures. One aspect of different structures is that some weave patterns are more open, permitting more light to pass trough. In terms of measurement, we characterize this as different levels of opacity. For these samples, the intrinsic reflectivity was calculated according to [16], where each substrate was measured first with just one layer on a black backing and after it was folded to an amount of textile thick enough to be completely opaque. Fig.3 shows the MCDM vs. opacity. For each value of opacity, the MCDM values are grouped on a small area, which can be a contribution

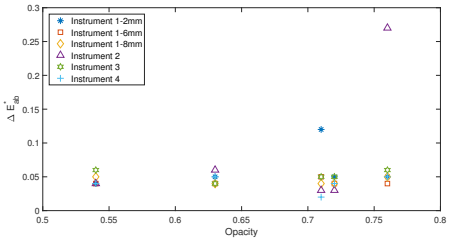


Fig. 3: MCDM against opacity for textile samples

TABLE IV: Inter-model agreement

		Inst. 1 2mm	Inst. 1 6mm	Inst.1 8mm	Inst. 2	Inst. 3	Inst. 4
Inst. 1-2mm	Avg ΔE^*_{ab}	-	0,17	0,17	2,85	1,06	1,94
	95%	-	0,23	0,26	3,43	2,40	3,39
Inst. 1-6mm	Avg ΔE^*_{ab}	0.17	-	0,14	2,79	0,99	1,89
	95%	0,23	-	0,19	3,36	2,40	3,41
Inst.1-8mm	Avg ΔE^*_{ab}	0,17	0,14	-	2,82	0,99	1,97
	95%	0,26	0,19	-	3,45	2,51	3,51
Inst. 2	Avg ΔE^*_{ab}	2,85	2,79	2,82	-	3,31	2,34
	95%	3,43	3,36	3,45	-	4,61	4,56
Inst. 3	Avg ΔE^*_{ab}	1,06	0,99	0,99	3,31	-	1,68
	95%	2,40	2,40	2,51	4,61	-	2,21
Inst. 4	Avg ΔE^*_{ab}	1,94	1,89	1,97	2,34	1,68	-
	95%	3,39	3,41	3,51	4,56	2,21	-

from the backing material, giving more consistency. On the element with higher opacity, the Instrument 2 has a different value than the other samples. This is related to the detection of fluorescent emission showed in Fig. 1a.

B. Inter-model agreement

Table IV shows the inter-model agreement between all instruments. The inter-model agreement between two instruments, according to [5] were calculated by equation (2), where I_A and I_B are two different instruments, s represents a sample and N the total number of samples. Since Instrument 4 has a different geometry, it is not expected to give compatible measurements with the 0°:45°devices, but the values are shown in Table IV for comparison. From all the comparisons, the best agreement is found between 6mm and 8mm from Instrument 1, which can is expected as only the aperture is varied. Also to be expected is that the poorest agreement is between instruments 2 and , i.e. between small-aperture 0°:45°and large-aperture 0°:d.

$$IMA(I_A, I_B) = \frac{1}{N}\sum_{i=1}^{N} \Delta E^*_{ab}(s_{i,A}, s_{i,B}) \qquad (2)$$

C. Measurement positioning variation on printed textiles

To characterize the effect of change in the spatial position of the measurement head with respect to the sample, a colour chart with eight patches digitally printed on the same textile samples described in Section II-B (cyan, magenta, yellow, black, red, green, blue and unprinted white) was measured.

Fig. 4 shows the measurement of the chart on textile sample 1 on CIELAB a*b* projection.

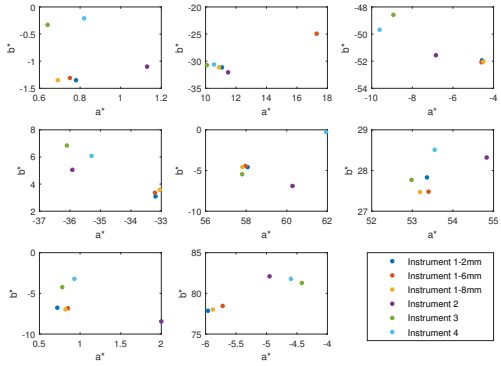
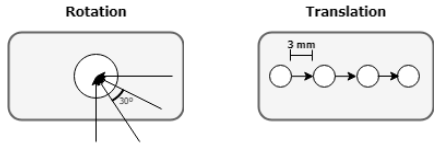


Fig. 4: Colour chart printed on textile sample 1 displayed on CIELAB a*,b* plane (from top left: black, blue, cyan, green, magenta, red, plain sample and yellow)

The substrates were measured using two different repositioning methods: a) rotation of the instrument with respect to the sample around the optical axis (see Fig. 5a); and b) horizontal translation of the instrument position by 3 mm with respect to the sample (Fig. 5b). The MCDM from the four measurements for rotation and translation respectively, for each textile sample was calculated. This was performed for each of the instruments described in Section II-A.

Fig. 6 shows the results for the measurements obtained by rotating the devices, with instruments ordered by the aperture. Instrument 1 using 2mm aperture and instrument 3 have the smallest apertures in this set of devices and also the highest ΔE^*_{ab} values, while the performance of 6mm and 8mm apertures depended on the sample (for the 45°:0°geometry). For samples 1 and 2, the 6mm was better and for samples 3, 4 and 5, 8mm had better results. The results are also shown in Table V.



(a) Representation of measurement rotating instruments

(b) Representation of measurement with horizontal translation

Fig. 5: Representation of how the samples were measured

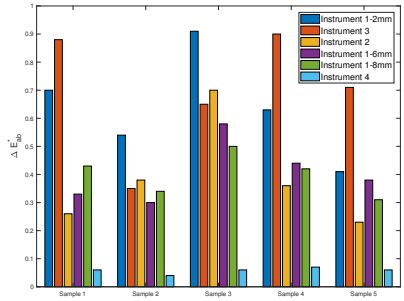


Fig. 6: Average ΔE^*_{ab} for measurements rotating the instruments ordered by the aperture size

Amongst the devices, instrument 4 as expected had better results overall owing to its 0°:d geometry, which minimises any differences caused by the rotation. The highest variation was found for Sample 3, possibly due to the fibres on this material being thinner and less opaque.

For the translation, Fig. 7 shows that the 2mm and 3mm apertures overall had lower repeatability, while the 8mm aperture instrument had similar or better results than 6mm. The measurements made with instrument 4 are substantially different from those shown in Fig. 6, where the variability is higher than in the rotation setup, and more similar to those of the other instruments. Comparing the difference between samples, just as the rotation, sample 3 also has the worse repeatability, supporting the idea that its weaving pattern leads to a higher degree of variability (See table VI).

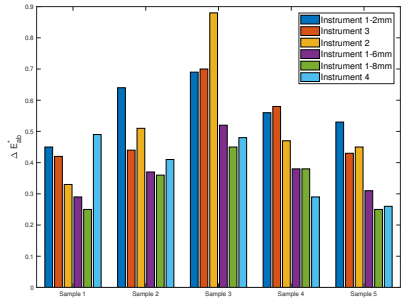


Fig. 7: Average ΔE^*_{ab} for measurements changing instrument's position ordered by the aperture size

In Figs. 8 and 9 it is possible to observe the effect the structure of the samples has on the measurements. Overall, there is a tendency of higher repeatability with increased sample opacity. This behavior is more evident with translation

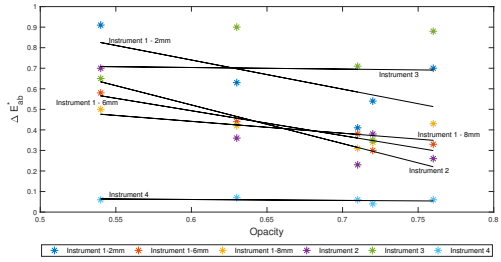


Fig. 8: Average ΔE^*_{ab} against opacity for measurements rotating instruments

TABLE V: MCDM and 95^{th} percentile for each sample rotating the instruments

		Inst. 1 2mm	Inst. 1 6mm	Inst.1 8mm	Inst. 2	Inst. 3	Inst. 4
Sample 1	MCDM	0.70	0.88	0.26	0.33	0.43	0.06
	95%	1.04	0.49	0.63	0.41	1.38	0.08
Sample 2	MCDM	0.54	0.35	0.38	0.30	0.34	0.04
	95%	0.80	0.50	0.53	0.57	0.52	0.07
Sample 3	MCDM	0.91	0.65	0.70	0.58	0.50	0.06
	95%	1.41	0.97	0.73	1.22	1.04	0.10
Sample 4	MCDM	0.63	0.90	0.36	0.44	0.42	0.07
	95%	0.86	0.63	0.62	0.61	1.45	0.10
Sample 5	MCDM	0.41	0.71	0.23	0.38	0.31	0.06
	95%	0.64	0.56	0.49	0.37	0.96	0.08

of the samples, since different regions of the same substrate are being compared. Figs. 10a and 10b show the average ΔE^*_{ab} of the samples for each aperture. In this set of instruments, only number 4 has a diffuse geometry; therefore, it has not been represented in these images.

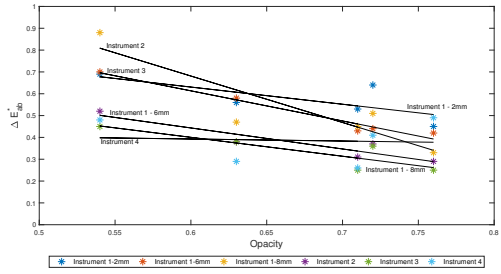
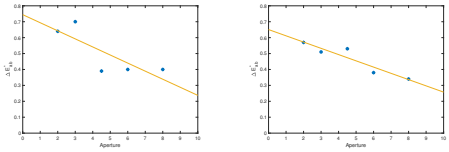


Fig. 9: Average ΔE^*_{ab} against opacity for measurements translating samples

TABLE VI: MCDM and 95^{th} percentile for each sample when measuring different areas

		Inst. 1 2mm	Inst. 1 6mm	Inst.1 8mm	Inst. 2	Inst. 3	Inst. 4
Sample 1	MCDM	0.45	0.42	0.33	0.29	0.25	0.49
	95%	0.73	0.42	0.38	0.52	0.62	0.95
Sample 2	MCDM	0.64	0.44	0.51	0.37	0.36	0.41
	95%	0.93	0.57	0.54	0.83	0.62	0.75
Sample 3	MCDM	0.69	0.70	0.88	0.52	0.45	0.48
	95%	1.08	0.81	0.74	1.39	1.01	0.88
Sample 4	MCDM	0.56	0.58	0.47	0.38	0.38	0.29
	95%	0.84	0.62	0.60	0.73	0.78	0.48
Sample 5	MCDM	0.53	0.43	0.45	0.31	0.25	0.26
	95%	0.73	0.42	0.38	0.52	0.62	0.95



(a) Average ΔE^*_{ab} against aperture for measurements rotating instruments (0°:45°geometry)

(b) Average ΔE^*_{ab} against aperture for measurements translating samples (0°:45°geometry)

Fig. 10: Average ΔE^*_{ab} against aperture

IV. Conclusion

We evaluated four different colour measurement instruments widely used in the graphic arts in order to understand the repeatability of these instruments on fabric materials used in textile printing.

On PTFE, all instruments performed consistently with the manufactures specifications. On textile samples, the short-term repeatability (as defined by MCDM) was in most cases of the same order as for the PTFE. When varying the position of the sample with respect to the instrument, the MCDM values were a little higher, but the performance was more dependent on both the instrument and the textile material.

The tests described indicate how for textile measurements the structure of the sample and the measurement aperture affect the results of color measurement. On printed samples, the instrument with diffuse geometry obtained the most reproducible results when using the device at the same location, regardless of the orientation to the sample. When measuring at different locations, the diffuse instrument has similar results to devices with a 45°:0°geometry. Hence we propose that directional instruments of the type used in the graphic arts industry can also be used in the control process of textiles, especially if used with a large aperture.

One instrument with three different options of aperture shows how the choice of aperture affects the measurement. Regardless of the sample structure, the larger the aperture the better the repeatability. When evaluating samples, the thread count alone is not sufficient to characterize textile materials. For example, sample 3 has more threads per cm^2 than sample 1; however, it has a lower opacity, resulting in larger ΔE^*_{ab} values. Hence, for digitally printed textiles, multiple characteristics of the substrate must be considered.

References

[1] ISO 13655:2017 (E), "Graphic technology- spectral measurement and colorimetric computation for graphic art images," Geneva, Switzerland, Tech. Rep. International Organization for Standardization, 2009.

[2] "Guide to the expression of uncertainty in measurement," Joint Committee for Guides in Metrology (JCGM/WG 1), Guide, 1995.

[3] ASTM E2214-08, "Standard practice for specifying and verifying the performance of color measuring instruments," Tech. Rep. American Society for the Testing of Materials (ASTM), 2008.

[4] ISO 105-J01:1997, "Textiles - tests for colour fastness - part j01: General principles for measurement of surface colour," Geneva, Switzerland, Tech. Rep. International Organization for Standardization, 2009.

[5] ISO/DTS 23031, "Graphic technology assessment and validation of the performance of spectrocolorimeters and spectrodensitometers," Geneva, Switzerland, Tech. Rep. International Organization for Standardization, 2018.

[6] J. Schanda, *Colorimetry: understanding the CIE system*. John Wiley & Sons, 2007.

[7] R. S. Berns *et al.*, *Billmeyer and Saltzman's principles of color technology*. Wiley New York, 2000.

[8] P. Nussbaum, A. Sole, and J. Y. Hardeberg, "Analysis of color measurement uncertainty in a color managed printing workflow," *Preliminary edition September 2011*, p. 7.

[9] D. R. Wyble and D. C. Rich, "Evaluation of methods for verifying the performance of color-measuring instruments. part i: Repeatability," *Color Research & Application*, vol. 32, no. 3, pp. 166–175, 2007.

[10] B. N. Taylor and C. E. Kuyatt, "Guidelines for evaluating and expressing the uncertainty of nist measurement results," National Institute of Standards and Technology (NIST), Tech. Rep. NIST Technical Note 1297, 1994.

[11] E. A. Early and M. E. Nadal, "Uncertainty analysis for reflectance colorimetry," *Color Research & Application*, vol. 29, no. 3, pp. 205–216, 2004.

[12] N. Milić, D. Novaković, and N. Kašiković, "Measurement uncertainty in colour characterization of printed textile materials," *Journal of Graphic Engineering and Design*, vol. 2, p. 2, 2011.

[13] "Textile sample presentation - an 1034.00," Application Note, HunterLab, 2015.

[14] "Color measurement methods for textile fabrics," White Paper, Datacolor, 2010.

[15] H. Gabrijelčič and K. Dimitrovski, "Influence of yarn count and warp and weft thread density on colour values of woven surface," *Fibres Text. East. Eur*, vol. 12, no. 1, p. 45, 2004.

[16] ISO 2471:2008 (E), "Paper and board determination of opacity (paper backing) diffuse reflectance method," Geneva, Switzerland, Tech. Rep. International Organization for Standardization, 2008.

Article G

Max Derhak, Phil Green, and Michele Conni. Color appearance processing using iccmax. *Electronic Imaging*, 2018(16):323–1, 2018

https://doi.org/10.2352/ISSN.2470-1173.2018.16.COLOR-323

Color appearance processing using iccMAX

Max Derhak, Onyx Graphics, Rochester, NY; Phil Green, NTNU in Gjøvik, Norway; Michele Conni, Barbieri Electronic, Bressanone, Italy

Abstract

ICC.2:2017 is a revision to the next-generation colour management specification iccMAX that introduces new support for colour appearance processing. iccMAX includes a built-in colour appearance model IccCAM, together with a rich programming environment, and support for spectral data, material channel connections, BRDF and processing elements that make it possible to functionally encode any appearance model. ICC.2:2017 introduces many new capabilities, including the ability to provide environment variables which allow parameters such as image statistics or viewing conditions to be passed to the transform at run-time. ICC.2:2017 supports a wide range of colour appearance computations within the colour management workflow.

Introduction

The ICC.1 profile specification [1], first published in 1996, established a well-defined architecture for communicating colour. ICC.1 was based on the concept of a reference intermediate colour space, and a profile which transformed all colour data encodings into or from this colour space, thus avoiding the combinatorial explosion of a many-to-many connection and enabling each colour device to be defined by a single profile.

This fixed Profile Connection Space (PCS), based on D50 colorimetry for a CIE 1931 2 degree standard observer, ensured interoperability of profiles regardless of who created them. The ICC.1 architecture employs a small set of transform elements – curve, matrix and multi-dimensional look-up table – that are applied in a predetermined sequence. The ICC.1 specification described the PCS for the Perceptual rendering intent as representing "the CIE colorimetry which will produce the desired color appearance if rendered on a reference imaging media and viewed in a reference viewing environment" [2]. This implies that as well as the device model the transform also embodies any gamut mapping, colour appearance and preference adjustments. In ICC.1 all these adjustments are combined into the values encoded in the AToB and BToA tags; using the curve, matrix and LUT transform elements there is no scope to encode the appearance transform separately, or to provide metadata on the transform or the conditions used, in a standard way.

The ICC.1 architecture also includes a default media-relative scaling of colorimetry. This can be over-ridden where needed, but a matching of source media white point to destination media white point is the most common expectation of colour management users. This media-relative scaling should not be confused with chromatic adaptation: although the form is similar to the Von Kries transform, both source and destination have a common illuminant (since all PCS colorimetry is required to be D50), and its function is primarily to ensure that source white is mapped to destination media white point. The adjustment made by media-relative scaling does in practice go some way to handling cross-media reproduction with different media white points where there **is** a degree of adaptation to the media white [3]. The media-relative PCS should be considered as a virtual colour space which allows two encodings to be connected, rather than a representation of actual colorimetry.

It can be seen from the above that ICC.1 is a well-defined but highly constrained architecture for colour transforms. Since 1996 many new requirements have emerged and the ICC.2 architecture [4], first published as an ICC specification in 2016, was designed to address them with a more flexible approach to connecting different colour spaces [5]. The ICC also publishes a Reference Implementation [6] which enables developers to make immediate use of ICC.2 constructs. The ICC.2 specification (also referred to as iccMAX) supports more flexible communication of:

i) Colorimetry. There is support for connection spaces other than D50, including specification of the illuminant and observer colour matching functions within the profile. Spectral data (reflectance or emission) is supported, both as input to a transform and as a PCS.

ii) Colour appearance. Colour appearance can readily be communicated by implementing a transform between different adapting conditions as a multiProcessElement within an iccMAX profile. For example, XYZ data for one adapting condition can be transformed via the desired appearance model into the adapting condition of the PCS of the profile. iccMAX incorporates a default IccCam model, which is a variant of CIECAM02.

iii) Other aspects of appearance. iccMAX includes directional appearance (through support for a number of BRDF models), and fluorescence (through support for input and processing of a full Donaldson matrix). Texture information can be communicated through height maps and normal maps associated with the BRDF model.

In addition to communicating appearance in terms of colour space values, iccMAX supports communication of metadata describing the adapting conditions as Spectral Viewing Conditions. The open format of the profile specification and the ability to convert between binary profiles and human-readable xml makes it possible for profile readers and users to extract the adapting viewing conditions from the profile and to modify them.

Finally, the ability to pass in an environment variable to parameterize the transform at the point when the profile is applied provides further flexibility. For example the environment variable can be used to input the adapting conditions at run-time, avoiding the need for an array of static profiles for each adapting condition.

The iccMAX profile format specifies a 32-bit floating point data type for all MPE transform elements, and both input and output data can also be encoded as 32-bit floats. The Reference Implementation source code can be compiled to perform computations at either 32-bit or 64-bit precision.

Application

In this paper we investigate some aspects of sensor adjustment transforms using iccMAX. A sensor adjustment transform (SAT) is used to transform colorimetry from one observing condition to a different observing condition based on various criteria [19]. Two types of sensor adjustment transform are considered in this paper:

chromatic adaptation transforms, and material adjustment transforms.

A chromatic adaptation transform (CAT) attempts to predict the corresponding colour for a given tristimulus value when the chromaticity of the adapting illumination changes. Chromatic adaptation has always been an important element of ICC.1 colour management, since in order to achieve interoperability all colorimetry with a different illuminant from the D50 PCS must be chromatically adapted to D50. ICC recommends a linearized version of the Bradford chromatic adaptation transform [1], which is implemented as a single 3x3 matrix. The matrix is stored in the ICC.1 profile and is used in the inverse direction when it is desired to transform from the PCS to the original colorimetry.

The CAT02 chromatic adaptation transform is an element of the CIECAM02 appearance model [8]. CIECAM02 performed well in predicting corresponding colour data sets, but a numerical instability in the transform has been reported and solutions proposed [9, 10, 11, 12, 13, 14, 15]. IccCam [4] replaces CAT02 by the HPE primaries and clipping to avoid negative RGB values. In the more recent CAT16, proposed as the chromatic adaptation transform in the CAM16 appearance model [15], the two stages of the CAT02 transform are replaced by a single matrix transform.

A common practice in implementing a CAT is to consider the inverse sense of the transform (where the test illuminant is the source and the reference illuminant is the destination) as the analytic inverse of the forward direction. Owing to the potentially different degree of adaptation in these two directions, and the tendency of the visual system to consider only near-daylight illuminants as neutral [18], it is suggested that where the test illuminant is chromatic the CAT model should not transform directly to it but using a two-step transform via a daylight illuminant or the equi-energy Illuminant E [15].

Chromatic adaptation transforms have generally been derived from corresponding-colour data sets and their performance evaluated in terms of their ability to predict such data [12]. Where the spectral reflectance for the colour exists, an alternative approach of forming a sensor adjustment transform to predicting the effect of a change of illuminant is of course to compute the tristimulus values for the test illuminant directly from the spectral data. This also suggests that a SAT can be derived from these XYZ values for reference and test illuminants computed from spectral reflectance, rather than via visual data sets. This approach is taken in e.g. [19, 20]. This type of transform assumes the reflectance of the object is unchanged and has been referred to as a Material Adjustment Transform (MAT) [19] or a Colorimetric Value Transform [20]. One distinguishing feature between a CAT and a MAT is that a material adjustment can provide a prediction of changes in observer in addition to changes in illuminant.

For connecting data encodings from different illuminants in a colour managed workflow, where both colorimetric and spectral data may be used, it is of interest to study the differences between the different SAT approaches in comparison to the tristimulus values under a different illuminant computed directly from spectral data. It has been shown that the degree of adaptation is reduced as the adapting illuminant becomes more chromatic and hence that CATs tend to over-predict the degree of adaptation and the resulting corresponding colours [16, 17].

Experimental

Several iccMAX profiles were used to transform colours from reflectance to colorimetry for the four adapting illuminants in Table 1. These colorimetric values were then transformed from the source (reference) illuminant to each of the other illuminants as destination or test illuminant, using the following SATs: CAT02 [8], CAT16 [15], linearized Bradford [1] and Wpt [19]. The degree of adaptation was set to 1.0 in all cases.

Table 1: Adapting illuminants

CIE illuminant	White point XYZ
D65	95.043, 100, 108.8801
D50	96.4197, 100, 82.5123
A	109.849, 100, 35.5825
F11	100.961, 100, 64.3506

The elements of the required adjustment transforms were computed in Matlab to determine the matrices used in the CustomToStandardPCS and StandardToCustomPCS tags. Using the existing iccMAX D65 colorimetric profile as a starting point, the xml was modified to encode these transforms, and profiles were subsequently created from the xml using the IccXml tool in the iccMAX Reference Implementation.

Two sets of reflectance data were selected. The first was the reflectance spectra of color chips from the Munsell Book of Color for Munsell Value 5 Chroma 6, measured at 1nm intervals [21] and subsequently corrected [22]. The second was the set of in-situ reflectance spectra from ISO 17231-1 [23]. The Wpt MAT was optimized for the corrected Munsell reflectances, so the ISO 17321 data set provides an independent test set.

The reflectance data were converted to XYZ for the reference illuminant using an iccMAX profile created to have a data colour space signature ‘nc0051’ representing 81 spectral channels, and a spectralViewingConditions tag populated by the CIE 1931 standard colorimetric observer over the range 380-780nm at 5nm intervals and the reference illuminant over the same range and interval. The command-line executable iccApplyNamedCmm [6] was called with the source data and profile as arguments. This was repeated for each of the reference illuminants in Table 1.

For each of the SATs tested, the transform was encoded in an iccMAX profile with an XYZ PCS and data colour space, in an A2B1 multiProcessElement-based tag, using the sequence of elements required by the transform. The XYZ data resulting from the previous step were converted to the test illuminant using iccApplyNamedCmm.

For each pair of reference and test illuminants, CAT02 and CAT16 transforms were performed both directly and via Illuminant E. The linearized Bradford transform assumes complete adaptation, so in CAT02 and CAT16 the degree of adaptation D was set to 1. The CAT02 and CAT16 transforms were also repeated with D = 0.93.

Results

The predicted XYZ values for each transform/illuminant combination were converted to CIELAB values, where the test illuminant was taken as the reference white in the conversion from XYZ to CIELAB. These values were compared with the XYZ values computed directly from Munsell and ISO 17321 in-situ reflectance spectra, similarly converted to CIELAB under the test illuminant. The test illuminant is expected to have CIELAB values of [100, 0, 0] after the above procedure. The differences between the

different methods are shown in Table 2 and Figures 3-9, for degree of adaptation $D = 1$.

Table 2 (A-D): Mean CIELAB ΔE^*_{ab} differences between XYZ values predicted from the reference illuminant chromatically adapted to the test illuminant, and the XYZ values for the test illuminant computed directly from the spectral reflectances

A: Reference illuminant D65

	Forward model			Via Illuminant E		
Munsell	**D50**	**A**	**F11**	**D50**	**A**	**F11**
CAT02	1.13	4.72	2.81	1.13	4.72	2.81
CAT16	1.40	6.91	3.15	1.39	6.91	3.17
Linearized Bradford	0.95	3.86	3.21	-	-	-
Wpt	0.55	1.84	2.23	-	-	-
ISO 17321	**D50**	**A**	**F11**	**D50**	**A**	**F11**
CAT02	1.79	6.19	5.37	1.79	6.19	5.37
CAT16	1.91	7.65	6.06	1.90	7.65	6.07
Linearized Bradford	1.66	5.54	5.95	-	-	-
Wpt	1.72	5.29	4.52	-	-	-

B: Reference illuminant D50

	Forward model			Via Illuminant E		
Munsell	**D65**	**A**	**F11**	**D65**	**A**	**F11**
CAT02	1.09	3.56	2.91	1.09	3.56	2.91
CAT16	1.36	4.45	2.87	1.36	4.48	2.87
Linearized Bradford	0.92	2.94	3.14	-	-	-
Wpt	0.56	1.33	1.83	-	-	-
ISO 17321	**D65**	**A**	**F11**	**D65**	**A**	**F11**
CAT02	1.78	4.51	5.92	1.78	4.51	5.92
CAT16	1.92	4.81	6.19	1.92	4.85	6.19
Linearized Bradford	1.67	4.04	6.20	-	-	-
Wpt	1.83	3.89	5.00	-	-	-

C: Reference illuminant F11

	Forward model			Via Illuminant E		
Munsell	**D65**	**D50**	**A**	**D65**	**D50**	**A**
CAT02	2.94	3.00	4.79	2.94	3.00	4.79
CAT16	3.28	2.95	5.37	3.28	2.96	5.38
Linearized Bradford	3.45	3.27	3.89	-	-	-
Wpt	2.10	1.81	1.63	-	-	-
ISO 17321	**D65**	**D50**	**A**	**D65**	**D50**	**A**
CAT02	5.60	6.10	8.33	5.60	6.10	8.33
CAT16	6.27	6.33	8.50	6.27	6.34	8.50
Linearized Bradford	6.29	6.43	7.22	-	-	-
Wpt	4.51	5.11	7.44	-	-	-

D: Reference illuminant A

	Forward model			Via Illuminant E		
Munsell	**D65**	**D50**	**F11**	**D65**	**D50**	**F11**
CAT02	4.15	3.27	4.61	4.15	3.27	4.61
CAT16	5.30	4.16	5.16	5.30	4.17	5.14
Linearized Bradford	2.94	3.72	3.89	-	-	-
Wpt	1.43	2.01	1.61	-	-	-
ISO 17321	**D65**	**D50**	**F11**	**D65**	**D50**	**F11**
CAT02	6.39	4.68	8.54	6.39	4.68	8.54
CAT16	6.5	4.72	8.45	6.50	4.70	8.40
Linearized Bradford	5.99	4.37	7.67	-	-	-
Wpt	7.23	5.01	8.52	-	-	-

The distribution of the differences in CIELAB ΔE^*_{ab} is shown in the histogram in Figures 1 and 2 for the Munsell and ISO 17321 data respectively.

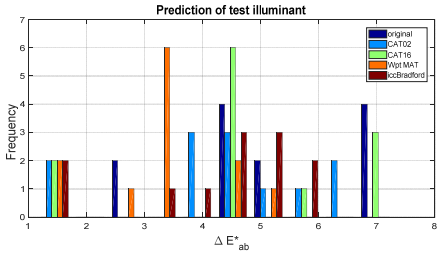


Figure 1. Distribution of mean differences between test illuminant calculated direct from spectra and predicted by SATs for all direct transform combinations in Table 1: Munsell and ISO 17321 in-situ reflectance data.

It can be seen from Table 2 that for the Munsell data set, Wpt-predicted XYZ values have smaller differences from the values computed for the test illuminant directly from the reflectances, compared with other transforms. However, when the ISO 17321-1 in-situ reflectances are considered Wpt has larger differences than the linearized Bradford transform.

It can also be seen from Table 1 and the examples in Figure 3 and 4 that for both CAT02 and CAT16, the differences between the predictions of the single-step transform and the transform via Illuminant E are negligible, as indicated in [15].

When the degree of adaptation was set to 0.93, the differences between the two workflows was similarly negligible, while the magnitude of difference was slightly larger when the reference illuminant was D50 and smaller when the reference illuminant was A.

In Figures 2-8 'original' represents the CIELAB a*, b* values of the colour computed from the reflectance under the reference illuminant; the CAT02 and CAT16 values are those predicted by the single-step transform from reference to test illuminant, and $CAT02_{23}$ and $CAT02_{24}$ values are transformed using the two-step workflow described in equations 23 and 24 in [15]. These can be compared with the values shown as 'from reflectance', which represent a*, b* values computed from reflectance under the test illuminant.

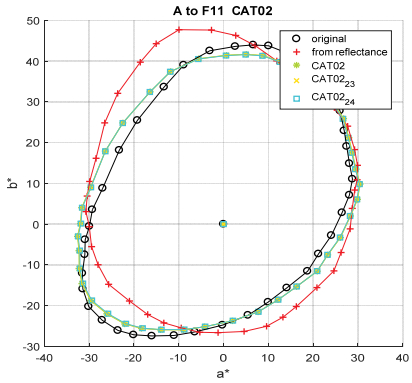


Figure 2. Differences in prediction of CAT02 using three workflows described in [15].

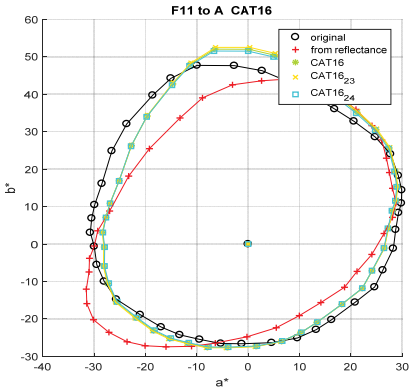


Figure 3. Differences in prediction of CAT16 using three workflows described in [15]

Examples of the differences between the predictions of the different SATs for the Munsell colours can be seen in Figures 4-8. As shown in Table 2, in most cases the Wpt transform tends to give the closest prediction of the Munsell colours computed from reflectance under the test illuminant.

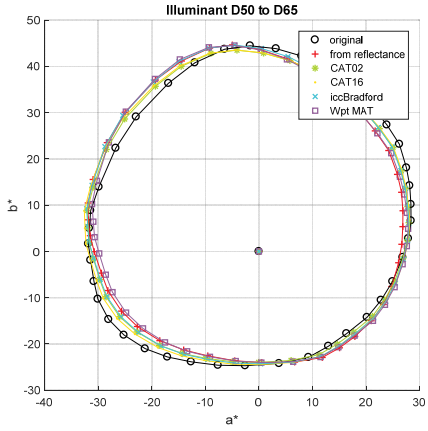


Figure 4. Munsell reflectances for D50 reference illuminant transformed to D65 by the 4 SATs

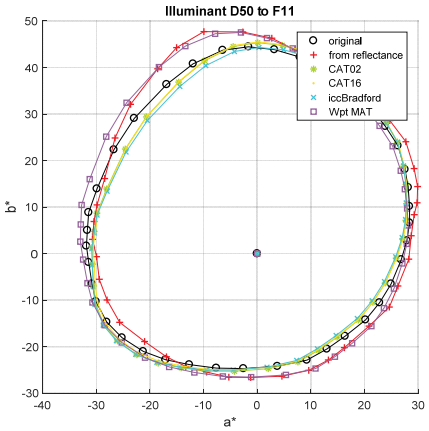


Figure 5. Munsell reflectances for D50 reference illuminant transformed to F11 by the 4 SATs

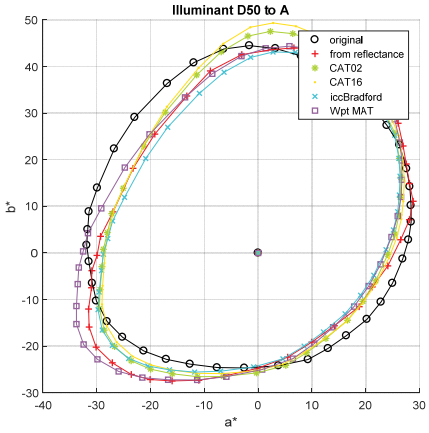


Figure 6. Munsell reflectances for D50 reference illuminant transformed to Illuminant A by the 4 SATs

As noted above, a CAT has a different derivation from a MAT and it is not necessarily expected that they should give equivalent results.

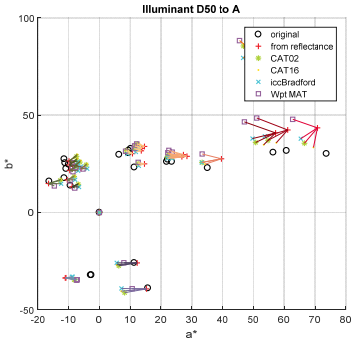


Figure 7. ISO 17321 in-situ reflectances for D65 reference illuminant predicted by the different SATs for D50, A and F11.

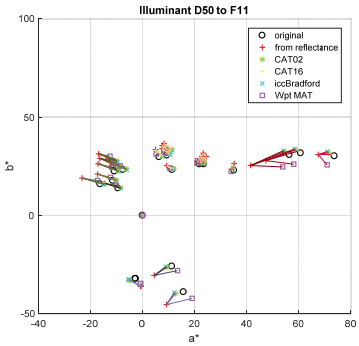


Figure 8. ISO 17321 in-situ reflectances for D65 reference illuminant predicted by the different SATs for D50, A and F11.

Conclusions

Four different transforms were used to predict the effect of a change in illuminant on two sets of reflectances, comprising the 1269 Munsell matt reflectances and the 24 ISO 17321-1 in-situ reflectances. Overall for the Munsell data the Wpt MAT gave the smallest differences between the tristimulus values predicted by the transform and those computed directly from the reflectance for the test illuminant, which is to be expected given that Wpt was optimized for the Munsell reflectances. CAT02 and CAT16 gave very similar predictions. For the ISO 17321-1 in-situ reflectance data the results vary with the reference and test illuminant and no single SAT performs best at predicting the colorimetry computed directly from reflectance, although the linearized Bradford transform adopted in ICC.1 performs reasonably well.

iccMAX provided a convenient framework for implementing the different transforms. Each transform was encoded as a v5 ICC profile using the IccFromXml tool in the iccMAX Reference Implementation. Conversion from reflectance to tristimulus values

for the source illuminant was performed using PCC profiles provided in the Reference Implementation, with the header modified in accordance with the wavelength range and interval used, and the conversion from source XYZ to destination XYZ for the different SATs was performed by DToB multiProcessingElements in profiles generated with IccXml. All CMM computations were performed at single precision using 32-bit floats.

The iccMAX framework also supports a wide range of other appearance processing elements. Using the Calc element it is possible to encode any appearance model. Although appearance coordinates are not available as a colour space data encoding in iccMAX, the PCS can be based on appearance coordinates making it possible to connect colour space data via transforms to and from appearance, and an Abstract class profile can be used to connect PCS to modified PCS coordinates. The Calc element also enables transform elements to be defined or selected at run-time.

Although the iccMAX framework provides a technical and computational framework for colour appearance processing, further work is needed to support the wider use of colour appearance models in colour management. Such activities include:

- Create and disseminate best practice recommendations for implementing appearance transforms in colour management applications.
- Develop publicly-available tools such as templates, source code, example profiles, and test data.
- Document the implementation of practical applications in iccMAX, including:

a) One or more colour appearance models

b) A seamless workflow to communicate colour appearance using ICC profiles

c) A workflow that supports parameterization of viewing conditions as environment variables

d) Interoperability Conformance Specifications for the above.

Acknowledgements

The authors wish to thank Professor M. R. Luo for his insightful comments.

References

[1] ISO 15076-1:2010 “Image technology colour management - Architecture, profile format, and data structure”

[2] ICC.1:2001-4 “File format for Color Profiles” http://www.color.org/icc_specs2.xalter

[3] K. Baah, P. Green and M. Pointer, “Perceived acceptability of colour matching for changing substrate white point”, Proc. SPIE 8652, Col. Imag. XVIII: Display, Processing, Hardcopy and Applications, 2013

[4] ICC.2:2017, “Image technology colour management - Extensions to architecture, profile format, and data structure (iccMAX)”

[5] M. Derhak, P. Green and T. Lianza “Introducing iccMAX - New Frontiers in Color Management”, Proc. SPIE 9395, Color Imaging XX: Display, Processing, Hardcopy, and Applications, 93950L, 2015

[6] ICC, “iccMAX Reference Implementation”, http://www.color.org/iccmax/, 2017

[7] R.W.G. Hunt, C. Li, and M.R. Luo, “Chromatic adaptation transforms”, Color Res. Appl., 30, 69-71, 2005

[8] N. Moroney, M. Fairchild, R.W.G. Hunt and C. Li, “The CIECAM02 color appearance model”, Proc IS&T: Tenth Color Imaging Conf.: Color Science and Engineering Systems, Technologies, Applications, 23-27, 2002

[9] C. Li and M.R. Luo, “Testing the robustness of CIECAM02”, Color Res Appl, 30, 99-106, 2005

[10] M. Brill and S. Süsstrunk, “Repairing gamut problems in CIECAM02: A progress report”, Color Res Appl 33, 424-426; and Erratum, p. 493, 2008

[11] C. Li, E. Perales, E., M.R. Luo, and F. Martínez-Verdú, “Mathematical approach for predicting non-negative tristimulus values using the CAT02 chromatic adaptation transform”, Color Res. Appl., 37, 255-260, 2012

[12] S. Süsstrunk, Sabine and G. Finlayson, “Evaluating Chromatic Adaptation Transform Performance”, Proc IS&T 13th Color and Imaging Conf., pp. 75-78(4), 2005

[13] M. Brill, “Irregularity in CIECAM02 and its avoidance”, Color Res. Appl., 31, 142-145, 2006

[14] M. Brill and M. Mahy, “Visualization of mathematical inconsistencies in CIECAM02”, Color Res. Appl., 38, 188-195, 2013

[15] C. Li, Z. Li, Z. Wang, Y. Xu, M.R. Luo, G, Cui, M. Melgosa, M. Brill and M. Pointer, “Comprehensive color solutions: CAM16, CAT16, and CAM16-UCS”, Color Res Appl. 42, 703-718, 2017

[16] K.A. Smet, Q. Zhai, M.R. Luo and P. Hanselaer, “Study of chromatic adaptation using memory color matches, Part II: colored illuminants”, Opt. Express 25, 8350-8365, 2017

[17] K.A. Smet, Q. Zhai, M.R. Luo and P. Hanselaer, “Study of chromatic adaptation using memory color matches, Part I: neutral illuminants”, Opt. Express 25, 7732-7748, 2017

[18] Q. Zhai and M.R. Luo, “A Study of Neutral White and Degree of Chromatic Adaptation”, Proc IS&T 25th Color Imaging Conf, Lillehammer, 2017

[19] M.W. Derhak and R.S. Berns, “Introducing Wpt (Waypoint): A color equivalency representation for defining a material adjustment transform”, Color Res. Appl., 40, 535-549, 2015

[20] Q. Liu, X. Wan, J. Liang, Z. Liu, D. Xie and C. Li, “Neural network approach to a colorimetric value transform based on a large-scale spectral dataset”. Coloration Technol, 133, 73-80, 2017

[21] U. Eastern Finland, “Munsell colors matt (spectrophotometer measured)” retrieved from http://www.uef.fi/web/spectral/munsell-colors-matt-spectrofotometer-measured

[22] M.W. Derhak and R.S. Berns, “Analysis and Correction of the Joensuu Munsell Glossy Spectral Database”. Proc. IS&T 20th Color Imaging Conf., 20: 191-194, 2012

[23] ISO 17321-1:2012 “Graphic technology and photography -- Colour characterisation of digital still cameras (DSCs) -- Part 1: Stimuli, metrology and test procedures”

Author Biography

Max Derhak has worked for Onyx Graphics Inc. since 1990 where he currently functions in the role of Principal Scientist. Max has a Bachelors in Computer Science from the University of Utah, a Masters in Imaging Science at The Rochester Institute of Technology, and a PhD. in Color Science from RIT. He serves as a Co-Chair of the ICC as well as the Chair of the ICC Architecture Working Group. He is also the initial contributor and maintainer of the iccMAX reference implementation - RefIccMAX.

Article H

Oswald Lanz, Fabian Sottsas, Michele Conni, Marco Boschetti, Erica Nocerino, Fabio Menna, and Fabio Remondino. A versatile multi-camera system for 3d acquisition and modeling. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:785–790, 2020

A VERSATILE MULTI-CAMERA SYSTEM FOR 3D ACQUISITION AND MODELING

O. Lanz * [1, 6], F. Sottsas [2], M. Conni [3], M. Boschetti [1, 4], E. Nocerino[5], F. Menna [6], F. Remondino [6]

[1] Covision Lab, Bressanone-Brixen, Italy - Email: <oswald.lanz><marco.boschetti>@covisionlab.com, Web: https://covisionlab.com

[2] Durst Phototechnik AG, Bressanone-Brixen, Italy

[3] Barbieri Electronics, Bressanone-Brixen, Italy

[4] Microtec, Bressanone-Brixen, Italy

[5] LIS UMR 7020, Aix-Marseille Université, CNRS, ENSAM, Université De Toulon, Marseille, 13397, France – erica.nocerino@univ-amu.fr

[6] Bruno Kessler Foundation (FBK), Trento, Italy – Email: <lanz><fmenna><remondino>@fbk.eu

Commission II – WG II/8

KEY WORDS: photogrammetry, computer vision, 3D modeling, point cloud, multi-view, system calibration

ABSTRACT:
Image-based 3D models generation typically involves three stages, namely: 2D image acquisition, data processing, and 3D surface generation and editing. The availability of different easy-to-use and low-cost image acquisition solutions, combined with open-source or commercial processing tools, has democratized the 3D reconstruction and digital twin generation. But high geometric and texture fidelity on small- to medium-scale objects as well as integrated commercial system for mass 3D digitization are not available. The paper presents our effort to build such a system, i.e. a market-ready multi-camera solution and a customized reconstruction process for mass 3D digitization of small to medium objects. The system is realized as a joint work between industrial and academic partners, in order to employ the latest technologies for the needs of the market. The proposed versatile image acquisition and processing system pushes to the limits the 3D digitization pipeline combining a rigid capturing system with photogrammetric reconstruction methods.

1. INTRODUCTION

Nowadays image- or range-based 3D reconstruction methods are receiving a lot of attention due to the availability of fast, easy-to-use and often low-cost hardware and software solutions. 3D has proven to be a promising approach to enable precise inspection, documentation, valorization, monitoring, communication, interaction and experience. 3D digital models, often called digital twins, are increasingly used in various fields and applications, such as e-commerce, website content, heritage restoration, industrial inspection and monitoring, digital archives and cataloguing, etc.

Digital twins provide an interactive browsing experience to users that can inspect digital items by zooming-in from any viewpoint. Many applications have strong demand on the fidelity of the reconstructed geometry and texture, especially if the 3D model has to be relighted when placed into virtual showrooms.

Even though 3D technologies and processing tools for small to medium objects have been democratized (Santos et al., 2017) and approaches for semantic enrichment (Grilli and Remondino, 2019; Pierdicca et al., 2020) and access to 3D models are starting to be used, few approaches enable mass digitization of a large variety of objects, from the heritage to the industry sectors.

Controlled acquisition setups are common in the 3D field for the digitization of small to medium scale objects. These setups can include only one camera or combine cameras, structured light and laser scanners. Objects can be placed on a moving turning table and recorded from fixed sensor positions in a programmed pose sequence (Santos et al., 2014; Gattet et al., 2015; Hosseininaveh et al., 2015; Menna et al., 2017). Open-source (Stathopoulou et al., 2019) or commercial 3D reconstruction tools can then be utilized to process the collected data and derive 3D surface models.

Although solutions based on motorized turntable are low-cost and easy to implement, they often do not meet industrial requirements and constraints. The main drawback of such systems is their limited speed of digitisation given the time needed to set up the system and allow the turntable to virtually move the sensor around the object according to an acquisition schema that places the object at the centre of a sphere. In these cases, the typical acquisition protocol requires a skilled operator to set up the sensor on a stand (i.e. a tripod) and adjust its position and orientation relatively to the object according to sensor's optical characteristics (working distance, depth of field, field of view, resolution) in order to meet the project tolerances. To accomplish a complete acquisition, the operator needs to move the sensor at different heights or tilt the object relatively to the sensor. Once the set-up of the system has been carried out, the time required to turn the table to acquire all the necessary data can be as high as several minutes for a single object. Moreover, the process is significantly influenced by the operator skills due to the required human intervention and interaction. For these reasons such solutions are not well suited for systematic digitization projects in industry where productivity and consistency in the produced results is of high importance. On the other hand, multi-camera setups, integrated with motorized turn tables and linear stages, despite intrinsic instrumentation costs, have proved great productivity (Santos et al., 2014) with a number of advantages such as that they can be accurately calibrated once in a while and allow fast data acquisitions. There are also many opportunities connected to multi-camera systems for high-quality 3D digitization of objects, in particular for the benefits provided in the post-processing and editing stages with less efforts by professionals required to match market needs.

* Corresponding author

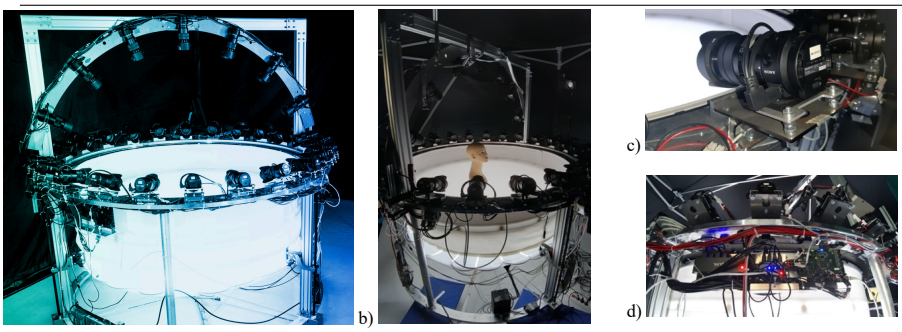


Figure 1: The DI-One multi-camera system with 31 PCB synchronized cameras, lights and a moving basement (a, b). One of the high-end professional cameras (c) and the electronics used to control groups of cameras (d).

1.1 Aim of the paper

In this paper we present our holistic approach for the realization of a market-ready multi-view camera system for mass 3D digitization of small to medium objects. The image acquisition and processing systems boosts the 3D digitization pipeline by jointly optimized photogrammetric reconstruction methods and a rigid capturing design and control. Compared to other solutions from the literature, the presented one (Figure 1) can deal with small up to medium objects (max size approximately 500x500x500 mm^3) thanks to an adaptable FOV, it has 31 synchronized high-end professional cameras tied on a rigid structure and an adjustable moving basement.
The rest of the paper is organized as follows: Section 2 describes the realized hardware system while Section 3 presents its geometric and radiometric calibration. Section 4 presents some experiments followed by conclusions in Section 5.

2. MULTI-CAMERA CAPTURING SYSTEM

The Durst Imaging Product System (DI-One - Figure 1) consists of 24+7 synchronized high-end professional cameras located on a spherical ring and on a hemispherical truss (half dome) to capture lateral and upper parts of objects. A directly and indirectly lighted semi-transparent and adjustable basement allows to place objects at different heights and inclination while the cameras can capture instantly 360 degrees images for different application fields and purposes. The system comes with a proprietary PCB designed to allow the system to trigger the cameras precisely via hardware with different trigger modes. Cameras can be triggered (i) all at the same time ("freeze / standard mode") or (ii) very accurately with a configured delay between each camera trigger ("sequential mode") or (iii) combining the two methods ("mixed mode"). The PCB design is modular. Besides, the camera trigger the PCB also controls additional equipment of the camera system like LED / light control or trigger of an attached flash. All acquired images (a "revolution") are transferred and written to disk each one in less than a second. A proprietary driver to control and trigger the cameras from a Linux system was developed, with a front-end application (GUI) that gives the user full control of the capturing system, i.e. change the settings of the cameras, set triggering modes, revolution history lookup, session controls, focus and zoom control, auto-focus and more.
The multi-view camera system has also a revolution playback system, i.e. a set of hardware and software components that listen for new files in a shared folder where newly created acquisitions are stored and plays them back as either a grid of images or a rapid slideshow, mimicking the playback of the 24-image revolution as a video.

3. SYSTEM CALIBRATION

The system calibration includes two procedures: (i) a geometric calibration (Section 3.1) to retrieve the camera parameters and (ii) a radiometric calibration (Section 3.2) to determine the correct colours in the images and improve the visual appearance of the generated 3D models.

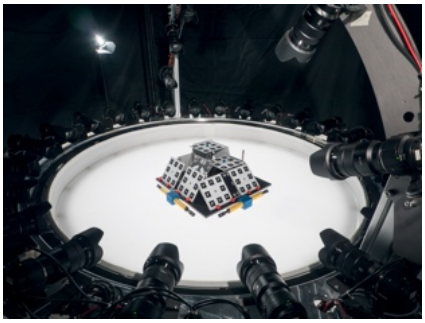
3.1 Geometric calibration

Photogrammetric self-calibration of a single or a multi-view camera system, based on ray intersection of multiple (target) points on a testfield or on a moved reference bar, was introduced by Brown (1971), Fraser (1997), Maas (1998), Gruen and Beyer (2001), Remondino and Fraser (2006).
The geometric calibration of the 31 cameras is performed using an ad-hoc testfield composed of scale bars and circular coded targets. The testfield (Figure 2) is placed in the centre of the basement in order to be visible by all cameras by rotating and tilting it in various positions. The testfield is modular and can be adjusted in size based on the necessary measurement volume. Targets of different diameters are used according to ground sample distance (GSD) and needed spatial resolution on the assets.
The calibration procedure is done in four steps:

1) a set of about 30 synchronized shots for all 31 cameras is acquired by tilting and rotating the testfield in the FOV of the cameras;
2) image orientation and bundle adjustment with self-calibration for each camera, using the 3D coordinates of the coded targets as soft constraints (coordinates are pre-measured with an estimated accuracy of better than 15 µm);
3) refinement of the exterior orientation parameters through a simultaneous bundle adjustment of all the acquired images (about 900) but keeping interior parameters fixed;
4) average relative orientation of all the cameras with respect to a master camera.

The calibration is repeated at different zoom levels, which can then be controlled via software. Each calibration is stored in the system's memory and utilized when needed (Fraser and Al-Ajlouni, 2006). The system accuracy and calibration verification over time is carried out using 3D length measurements of invar scale bars (Figure 2-bottom). The measurements obtained with

the DI-One system are checked against reference calibrated values (VDI/VDE, 2002).



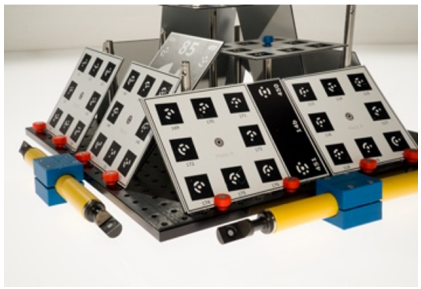


Figure 2: The used testfields for geometric calibration of the camera system (interiors and poses).

Figure 3 shows the complete image network used in the bundle adjustment (ca 900 images) together with the 3D coordinates of the testfield.

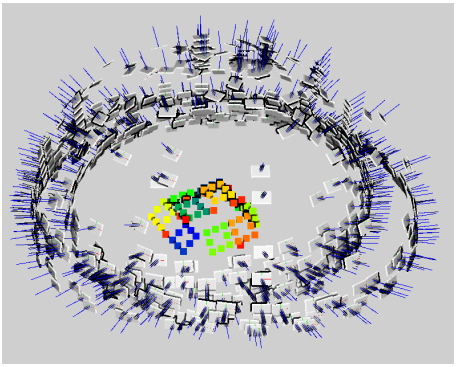


Figure 3: example of a full camera network used for the geometric calibration.

3.2 Radiometric calibration

The DI-One system delivers multi-camera shots in RAW image format which need to be radiometrically calibrated and post-processed to meet the colour fidelity requirements of professional applications. Assuming that the object surfaces to be acquired are Lambertian, the acquisition process can be expressed through the equation (Gaiani et al., 2017):

$$f_c(x) = m(x) \int_{\omega} I(\lambda)\rho(\lambda)S(x,\lambda)d\lambda \qquad \text{Eq. (1)}$$

where:
x = spatial coordinate
λ = wavelength
ω = camera's spectral domain (visible)
c = colour index (R, G, B)
$f_c(x)$ = raw value of the image at position x, filter colour c
$m(x)$ = Lambertian shading
$I(\lambda)$ = radiant intensity of the light source
$\rho_c(\lambda)$ = camera sensitivity function for colour c
$S(x, \lambda)$ = spectral reflectance of the surface.

Equation 1 shows the dependence of a camera's pixel response on its sensitivity function $\rho_c(\lambda)$ which generally varies from camera to camera (Pagnutti et al., 2017).
$f_c(x)$ is the raw value at position x of an image acquired with a commercial camera, and in a commercial imaging pipeline it undergoes multiple steps of post-processing, such as Bayer interpolation, noise subtraction, and white balancing.
If a faithful chromatic reproduction of an object is required, these values, which represent the camera's own RGB values, must be linked to a device-independent colour space, such as CIEXYZ or CIELAB. This process is referred to as the radiometric camera's calibration (Westland et al., 2012) and it can be performed in various ways.
The most widely used characterization approaches are described in the international standard ISO 17321-1 (ISO17321): one is based on the measurement of the cameras' spectral sensitivities, while the other is target-based.
We performed the latter procedure for each camera in the multi-camera system. Since the illumination is controlled, the colour calibration was performed for a D50 standard illuminant.
We printed an IT8.7/4 colour chart with 1617 patches on a ProofMaster Matt 140g paper (Fig. 4). Although such a target has been designed for the colour characterisation of scanners, monitors and other output devices, it has been previously exploited to perform multi-camera calibration (Troester et al, 2018).

Figure 4: Example of captured IT8.7/4 colour chart

We then measured the CIELAB coordinates of the target with a Barbieri Electronic LFPqb spectro-photometer (Barbieri, 2020), and we used them as references.
The colour chart was taped on a flat white matte holder and was acquired with each camera of the system at a position roughly normal to its optical axis. A spotlight with 5000 K colour

temperature was placed at about 45 degrees with respect to the camera axis. The other acquisition parameters correspond to those previously discussed.

On each image we applied the Python's module rawpy (https://pypi.org/project/rawpy) for an AHD demosaicing and for the extraction of the raw RGB values of each colour patch. It is then possible to link these values to the previously measured CIELAB values in order to obtain a proper colour profile, which can then be used to correct the texture file of a 3D model.

In particular, we used a multi-dimensional look-up table defined on a 33x33x33 grid in the raw RGB space of each camera (Balasubramanian and Klassen, 2003), implemented using the Python's scikit-learn tool (https://scikit-learn.org/).

An example of the radiometric processing is shown in Fig. 5. The improvement in the appearance of the scene is evident. If necessary, further corrections could be performed, such as bad pixel removal, bias and dark frame subtraction, flat fielding, etc.

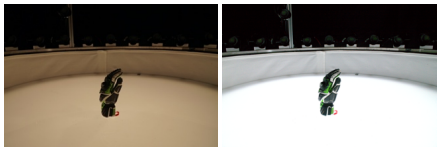


Figure 5: Example of the radiometric calibration result (right) on a raw image (left) of a glove.

4. 3D RECONSTRUCTION PIPELINE

The DI-One system (https://www.covisionlab.com/media-lab) is coupled with an automated photogrammetric pipeline which allows to generate textured 3D models of various objects (Figure 6). Using the camera calibration parameters, the images are remapped as if they were all obtained through the same pinhole camera, i.e. without geometric distortions, with a unique principal distance (focal length) and the principal point at the center of the image format. This procedure simplifies the 3D reconstruction and modelling pipeline and allows the use of the DI-One images also with other research and commercial based software, which might not be able to deal with distortion parameters or use a different distortion formulation.

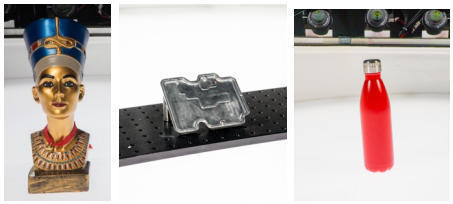


Figure 6: Examples of some objects digitized with the realized DI-ONE multi-camera system.

For objects characterised by cooperative texture and simple geometry, a single DI-One synchronized shot (31 images) may be sufficient for reconstructing their visible parts. In these cases, the exterior orientation parameters (camera poses) obtained from system calibration (Section 3.1) are directly used. Figure 7 shows the coordinate system and the camera network as defined from the calibration stage. These exterior orientation parameters are directly provided to the photogrammetric pipeline for the 3D reconstruction procedure, thus providing also the proper scaling.

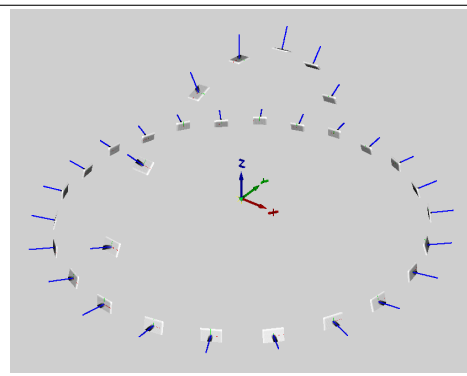


Figure 7: DI-One camera network and coordinate system.

The typical pipeline starts with a background masking, performed automatically through image subtraction procedures (Figure 8) and then feature points are extracted and matched within the unmasked areas of the images. These tie points are then triangulated using the exterior orientation parameters and the individual interior orientation parameters derived from the calibration procedure. Dense image matching procedures are then applied to generate depth maps and a dense cloud. Finally, a mesh is triangulated over the point cloud and texturized.

Figure 8: Example of image masking obtained through background subtraction.

Figure 9 shows an example of a 3D reconstruction of small vase (170x130x180 mm^3) obtained with 31 synchronized images acquired under the camera network shown in Figure 7.

Figure 9: One (out of 31) synchronized image of a small vase obtained with the DI-One system (left) and the reconstructed 3D mesh (right).

For more complex objects (such as the glove in Figure 12), depending on their shape, self-occlusions and texture characteristics (shiny, texture less, etc), more shots may be required after tilting or rotating the object within the field of view of the system. Different camera networks can thus be obtained in order to achieve a complete 3D reconstruction. Figures 10 and 11 show two examples of camera networks obtained with a total of 4 DI-One shots by rotating the object respectively around the Z

axis by the angles {-48.75, 48.75, 97.5} [deg] (Figure 10) or by tilting the object around the Y axis of the angles {-22.5, 22.5, 45} [deg] (Figure 11).

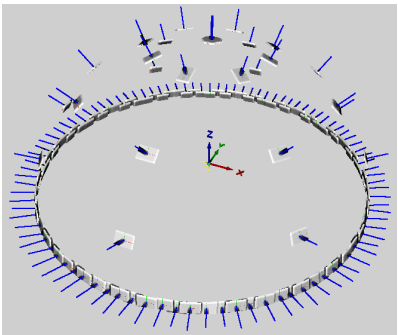


Figure 10: A typical camera network obtained rotating the object around the Z axis in order to have a better coverage.

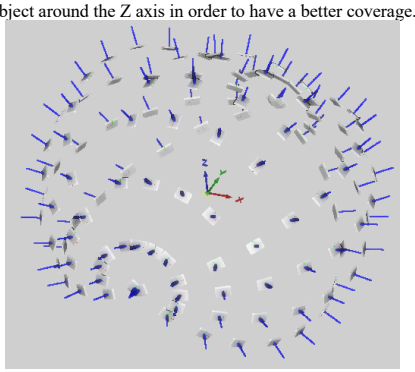


Figure 11: Examples of camera network obtained using multiple shots after tilting the object around the Y axes.

Figure 12 shows two rendered views of the 3D texturized model of a Durst Gil camera (production 1938), reconstructed from two shots to capture the entire object texture and geometry.

Figure 12: Views of a 3D model of the 1938 *Durst Gil* camera reconstructed from two shots taken at different object poses.

Figure 13 shows the 3D reconstruction of a skiing glove pictured with three synchronized shots (93 images) and textured with a radiometrically corrected texture.

Figure 13: One (out of 93) image of a glove obtained by rotating the system basement around the Z axis (left) and the reconstructed 3D mesh through photogrammetric process (right).

5. CONCLUSIONS

The paper presented the realization of the Covision Media Lab (https://www.covisionlab.com/media-lab) 3D digitization process named DI-One multi-camera system. DI-One is a market-ready image-based solution composed of 31 synchronized high-end professional cameras for mass 3D digitization of small to medium objects. The system and processing methodology were conceived as a joint work between industrial and academic partners, they are quite versatile and push to the limits the 3D digitization by combining a rigid capturing system with photogrammetric reconstruction methods.

ACKNOWLEDGEMENTS

The authors would like to highlight the contribution of Durst Phototechnik AG in this effort to manufacture the Product System DI-One.

REFERENCES

Balasubramanian, R., Klassen, R. V., 2003. Efficient color transformation implementation. *Digital Color Imaging Handbook*, CRC Press: Boca Raton, FL, USA.

Barbieri, 2000. https://www.barbierielectronic.com/en/digital-imaging/1-0.html (last access: May, 2020).

Brown, D., 1971. Close-Range Camera Calibration. *Photogrammetric Engineering*, Vol. 37, No. 8, pp. 855-866.

Fraser, C.S. and Al-Ajlouni, S., 2006. Zoom-dependent camera calibration in digital close-range photogrammetry. *Photogrammetric Engineering & Remote Sensing*, 72(9), pp.1017-1026.

Fraser, C.S., 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 52, pp. 149-159.

Gaiani, M., Apollonio, F. I., Ballabeni, A., Remondino, F., 2017. Securing color fidelity in 3D architectural heritage scenarios. *Sensors*, 17(11), 2437.

Gattet, E., Devogelaere, J., Raffin, R., Bergerot, L., Daniel, M., Jockey, Ph., De Luca, L., 2015. A versatile and low-cost 3D acquisition and processing pipeline for collecting mass of

archaeological findings on the field. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XL-5/W4.

Grilli, E., Remondino, F., 2019. Classification of 3D Digital Heritage. *Remote Sensing*, Vol. 11(7), 847.

Gruen, A. and Beyer, H.A., 2001. System calibration through self-calibration. In '*Calibration and Orientation of Cameras in Computer Vision*' Gruen and Huang (Eds.), Springer Series in Information Sciences 34, pp. 163-194.

Hosseininaveh, A., Yazdan, R., Karami, A., Moradi, M. and Ghorbani, F., 2015. A low-cost and portable system for 3D reconstruction of texture-less objects. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 40(1/W5), p.327.

ISO 17321-1:2012, 2012. Graphic technology and photography—Colour characterisation of digital still cameras (DSCs) - Part 1: Stimuli, metrology and test procedures.

Menna, F., Nocerino, E., Morabito, D., Farella, E. M., Perini, M., Remondino, 2017. An open source low-cost automatic system for image-based 3D digitization. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. XLII-2/W8, pp. 155-162.

Pagnutti, M. A., Ryan, R. E., Cazenavette, G. J., Gold, M. J., Harlan, R., Leggett, E., Pagnutti, J. F., 2017. Laying the foundation to use Raspberry Pi 3 V2 camera module imagery for scientific and engineering purposes. *Journal of Electronic Imaging*, Vol. 26(1), 013014.

Pierdicca, R., Paolanti, M., Matrone, F., Martini, M., Morbidoni, C., Malinverni, E.S., Frontoni, E., Lingua, A.M., 2000. Point Cloud semantic segmentation using a deep learning framework for Cultural Heritage. *Remote Sensing*, 12, 1005.

Remondino, F., Fraser, C., 2006. Digital camera calibration methods: considerations and comparisons. *ISPRS Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Vol. 36(5), pp. 266-272.

Santos, P., Ritz, M., Tausch, R., Schmedt, H., Monroy, R., De Stefano, A., Posniak, O., Fuhrmann, C., Fellner, D.W., 2014. CultLab3D - On the verge of 3D mass digitization. Proc. *Eurographics Workshop on Graphics and Cultural Heritage*.

Santos, P., Ritz, M., Fuhrmann, C. and Fellner, D., 2017. 3D mass digitization: a milestone for archaeological documentation. *Virtual Archaeology Review*, Vol. 8(1).

Stathopoulou, E.-K., Welponer, M., and Remondino, F., 2019. Open-source image-based 3D reconstruction pipelines: review, comparison and evaluation. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W17, 331-338.

Troester, P., Nestler, R., Beger, M., Urban, P., 2018. Fogra-Forschungsbericht Nr. 11.002 Prozessstandardisierung im grafischen 3D-Druck fuer 3DP- und FDM-Verfahren, Munich, Germany. Available at: https://fogra.org/en/fogra-research/prepress/research-projects-prepress/graphical-3d-printing/publications-3d.html (last access: May, 2020).

VDI/VDE, 2002. Guideline 2634 part 1, Optical 3D measuring systems - Imaging systems with point-by-point probing. VDI/VDE-Gesellschaft Mess-und Automatisierungstechnik.

Westland, S., Ripamonti, C., Cheung, V., 2012. *Computational colour science using MATLAB*. John Wiley & Sons.

Oral contributions

Michele Conni. iccMAX in Barbieri. https://www.color.org/DevCon/devcon2020, 2020. Presented at ICC DevCon 2020: The Future of Color Management

Michele Conni
R&D Engineer at Barbieri Electronic
Ph.D. candidate at NTNU

Outline

1. Barbieri
2. What is texture?
3. Texture models
4. Roughness correction in iccMAX

1. Barbieri

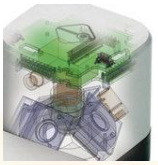
Products

Barbieri spectrophotometers

Spectro LFP qb for Large Format Printing

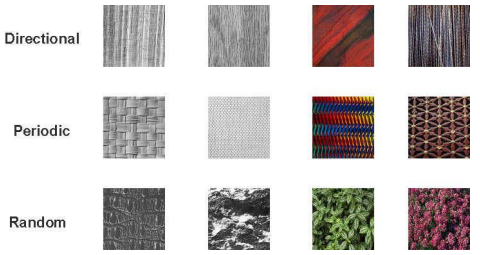
Spectro Pad as portable solution for roll-to-roll Format Printing

Spectro Swing for calibration in roll-to-roll-Format Printing



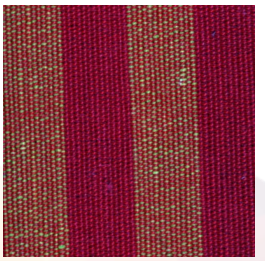
Customized measuring devices for OEM-manufacturers

2. What is texture?



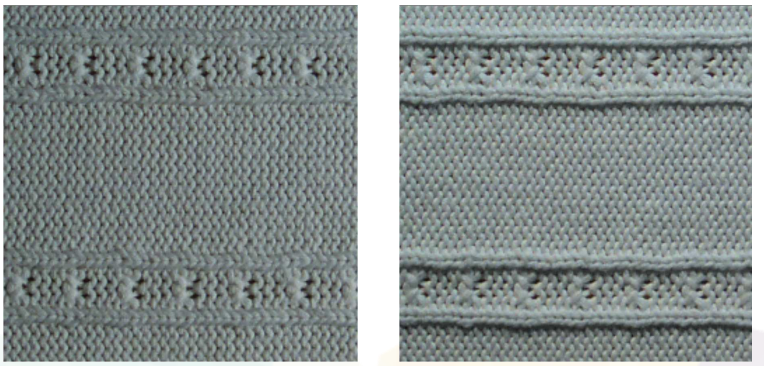
- "No formal definition of texture exists, intuitively this descriptor provides measures of properties such as smoothness, coarseness and regularity." [Gonzalez, 2002]

- Usually refers to a scene taken from a single object/material characterized by spatial complexity



ICC DevCon 2020 © MICHELE CONNI 5

Surface texture

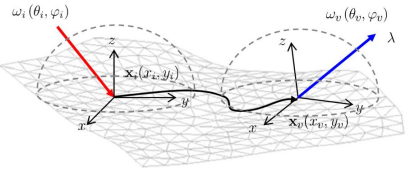


From [Dong, 2005]

ICC DevCon 2020 © MICHELE CONNI 6

3. Texture models

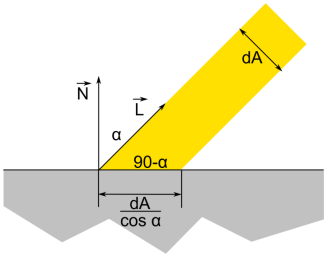
- General Reflectance Function (GRF): 16 variables source, detector, collision, emission coordinates + time and frequency of generation and detection
- Bidirectional Surface Scattering Reflectance Distribution Function (BSSRDF): 9D (scattering)
- Bidirectional Texture Function (BTF): 7D (surface)
- Bidirectional Reflectance Distribution Function (BRDF): 5D (point)
- Drawbacks: lengthy and expensive processes, cumbersome data management



From [Haindl, 2013]

Lambertian reflectance model

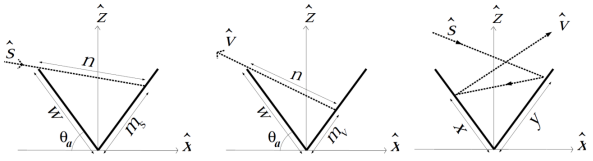
- Property of ideal diffusely reflecting surface
- Surface reflectance is isotropic
- It is impossible to tell where the incident light comes from
- No specular peak
- Real world examples: matte paper, flat paint, opal glass

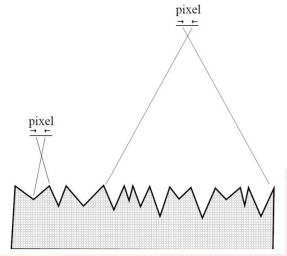


From [wikipedia]

Oren–Nayar reflectance model

- Comprehensive model for body reflectance from surfaces with macroscopic roughness
- Accounts for complex geometric and radiometric phenomena (masking, shadowing, interreflections)
- Based on V-cavities
- Depends on the acquisition system (e.g. resolution of pixels)



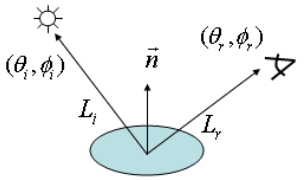


From [Oren, 1994]

Oren–Nayar reflectance model

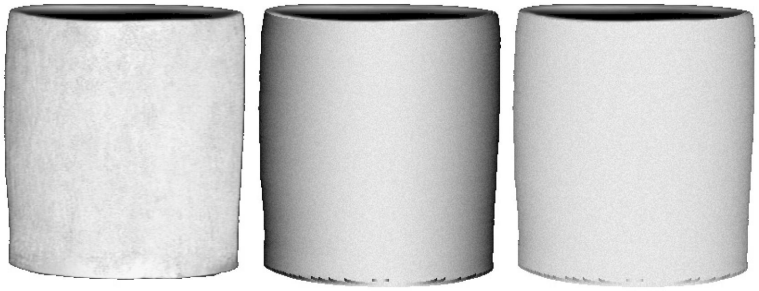
$$L_r = \frac{\rho}{\pi}\cos\theta_i[A(\sigma) + B(\sigma)\max(0,\cos(\phi_i - \phi_r))\sin(\max(\theta_i,\theta_r))\tan(\min(\theta_i,\theta_r)))]\,E_0$$

- Statistical model
- Effective for rough diffuse surfaces, such as, plaster, sand, clay, and cloth
- $A(\sigma) = 1 - 0.5\frac{\sigma^2}{\sigma^2+0.33}$
- $B(\sigma) = 0.45\frac{\sigma^2}{\sigma^2+0.09}$



From [wikipedia]

Oren–Nayar reflectance model

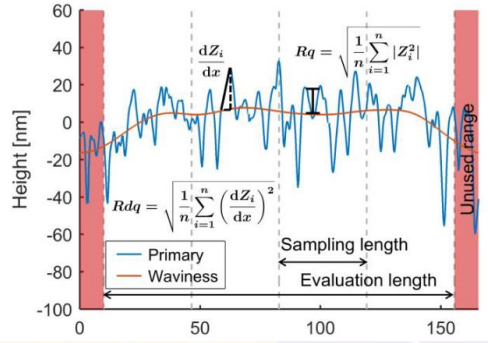


(a) Image (b) Lambertian (c) Model

From [Oren, 1994]

Roughness

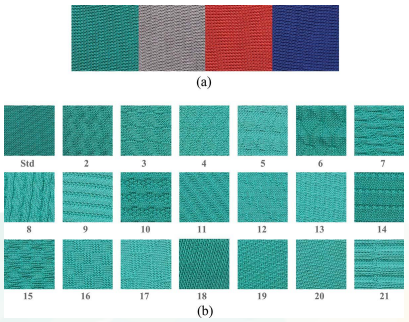
- More models exist
- E.g., Principled BRDF
- $Rdq = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{dZ}{\mathrm{dx}}\right)_i^2}$
- Lambertian and Oren-Nayar models can be mixed

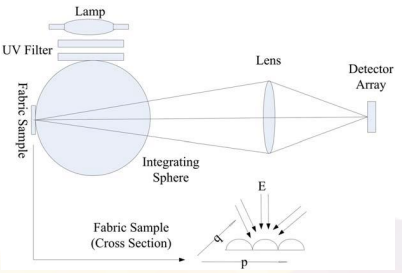


From [Feidenhans'l, 2015]

Luo texture correction model

- Use reflectance model to correct colour measurements of textiles
- The correction must be done in CIEXYZ, given linearity with reflectance
- $X = \frac{\iint_{p,q} m_b(p,q)H(p,q)dpdq}{A_r} \int_\lambda E(\lambda)R(\lambda)\bar{x}(\lambda)d\lambda = C \cdot X_{norm}$

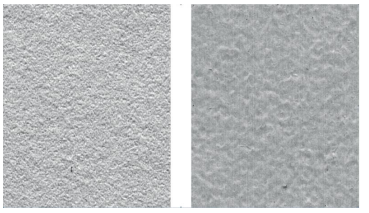




From [Luo, 2014]

4. Roughness correction in iccMAX

- Extension of ICC v4
- v5 in header
- Backwards compatibility
- ISO 20677
- Novelties examples:
- Extended connection space (e.g. flexible PCS, illuminant, CMFs)
- multiProcessingElements (matrices, LUTs, CAM and Calc elements)
- Spectral and BRDF support
- Height/normal map can be stored but not used in profile
- Can be used for rendering



From [Specification ICC.2:2019]

The Calc element

- MultiProcessElement
- Stack-based programming language
- Uses reverse polish notation: $2 * 4 \rightarrow 2\ 4\ mul$
- Structure of CalculatorElement:

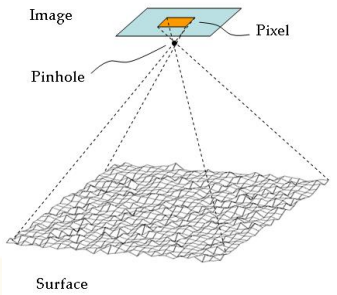
```
<CalculatorElement InputChannels="in" OutputChannels="out" InputNames="…" OutputNames="…">
        <Imports> … </Imports>
        <Variables> … </Variables>
        <Macros> … </Macros>
        <SubElements>… </SubElements>

        <MainFunction>Extended Textual Representation of Operations</MainFunction>
</CalculatorElement>
```

- The data is private to the MultiProcessElement
- Stack empty at the start

Implementation

- Two texture correction models: Lambertian and Oren–Nayar
- Purpose: correction of colour measurement on complex surface
- $XYZ_{norm} = XYZ_{meas}/C$
- $C_{Lambert} = \mu(\cos(\theta_i))$
- $C_{Oren-Nayar} = \mu(\cos\theta_i[A(\sigma) + B(\sigma)\max(0, \cos(\phi_i - \phi_r))\sin(\max(\theta_i, \theta_r))\tan(\min(\theta_i, \theta_r)))])$
- Single profile, CIEXYZ data colour space, CIEXYZ PCS
- Relative Colorimetric rendering intent
- ColorSpace ('spac') profile
- Correction implemented in multiProcessElement, A2B1 tag
- B2A1 tag implements $XYZ_{meas} = C \cdot XYZ_{norm}$



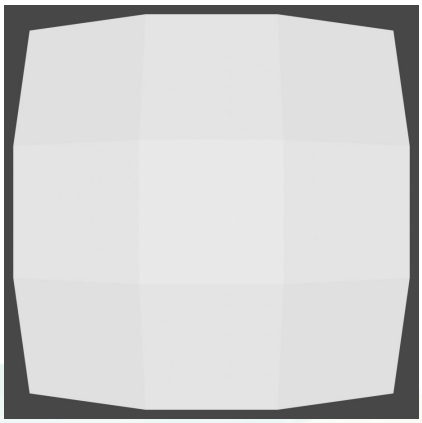
From [wikipedia]

Lambertian model correction

- $XYZ_{norm} = XYZ_{meas}/\mu(\cos(\theta_i))$
- Declarations:

```
<multiProcessElementType>
    <TagSignature>A2B1</TagSignature>
    <MultiProcessElements InputChannels="3" OutputChannels="3">
        <CalculatorElement InputChannels="3" OutputChannels="3">
            <Imports>
                <Import Filename="import_lamb.xml"/>
            </Imports>
            <Variables>
                <Declare Name="myVector" Size="9"/>
            </Variables>
            <Macros>
                <!-- Macro to convert degrees to radians -->
                <Macro Name="torad">3.14159265359 mul 180 div</Macro>
            </Macros>
        <MainFunction>
```


3D structure



Front

Side

Lambertian model correction

- import_lamb.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<IccCalcImport>
    <Macros>
        <!-- Map of polar angles of incident light with each facelet -->
        <Macro Name="ti">
            44.14 44.14 44.14
            55.00 45.00 35.00
            44.14 44.14 44.14
        </Macro>
    </Macros>
</IccCalcImport>
```


Lambertian model correction

- Main function:

```
        <MainFunction>
{
in(0,3)
% Load normal map
call{ti}
% Calculate cosine and save in myVector
call{torad} cos tput{myVector[0]}
call{torad} cos tput{myVector[1]}
call{torad} cos tput{myVector[2]}
call{torad} cos tput{myVector[3]}
call{torad} cos tput{myVector[4]}
call{torad} cos tput{myVector[5]}
call{torad} cos tput{myVector[6]}
call{torad} cos tput{myVector[7]}
call{torad} cos tput{myVector[8]}
% Calculate average cosine
tget{myVector} sum(9)
9 div
% Use it to correct XYZ
copy copy
div(3)
out(0,3)
}
        </MainFunction>
```

Oren-Nayar model correction

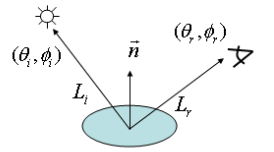
- $XYZ_{norm} = XYZ_{meas}/\mu\,(\cos\theta_i[A(\sigma) + B(\sigma)\max(0, \cos(\phi_i - \phi_r))\sin(\max(\theta_i, \theta_r))\tan(\min(\theta_i, \theta_r)))])$
- Declarations:

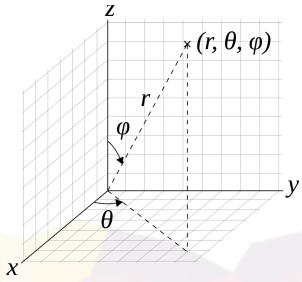
```
<multiProcessElementType>
    <TagSignature>A2B1</TagSignature>
    <MultiProcessElements InputChannels="3" OutputChannels="3">
        <CalculatorElement InputChannels="3" OutputChannels="3">
            <Imports>
                <Import Filename="import_oren.xml"/>
            </Imports>
            <Variables>
                <Declare Name="cos_ti" Size="9"/>
                <Declare Name="A"/>
                <Declare Name="B"/>
                <Declare Name="phi_i" Size="9"/>
                <Declare Name="phi_r" Size="9"/>
                <Declare Name="t_i" Size="9"/>
                <Declare Name="t_r" Size="9"/>
                <Declare Name="cos_diff" Size="9"/>
                <Declare Name="sin_alpha" Size="9"/>
                <Declare Name="tan_beta" Size="9"/>
            </Variables>
```


Oren-Nayar model correction

- import_oren.xml:

```
<?xml version="1.0" encoding="UTF-8"?>
<IccCalcImport>
    <Macros>
        <!-- Here are stored the parameters of the Oren-Nayar correction model -->
        <!-- Std dev of the model -->
        <Macro Name="sigma">
            0.50
        </Macro>
        <!-- Map of polar angles of incident light with each facelet -->
        <Macro Name="ti">
            44.14 44.14 44.14
            55.00 45.00 35.00
            44.14 44.14 44.14
        </Macro>
        <!-- Map of azimuth angles of incident light with each facelet -->
        <Macro Name="phii">
            -125.15   9.85  54.85
              0.00    0.00   0.00
            144.85 350.15 -35.15
        </Macro>
        <!-- Map of polar angles of detector with each facelet -->
        <Macro Name="tr">
            80.00 10.00 80.00
            10.00  0.00 10.00
            80.00 10.00 80.00
        </Macro>
        <!-- Map of azimuth angles of detector with each facelet -->
        <Macro Name="phir">
             -45.00  90.00 135.00
               0.00   0.00 180.00
            -135.00 270.00  45.00
        </Macro>
    </Macros>
</IccCalcImport>
```





From [wikipedia]

Oren-Nayar model correction

- Conversion to radiants and $\cos(\theta_i)$:

```
<Macros>
    <!-- Macro to convert degrees to radians -->
    <Macro Name="torad">3.14159265359 mul 180 div</Macro>
    <Macro Name="calc_cos_ti">
        % Load illumination cosines map
        call{ti} tput{t_i}
        % Calculate cosine and save in cos_ti
        tget{t_i[0]} call{torad} cos tput{cos_ti[0]}
        tget{t_i[1]} call{torad} cos tput{cos_ti[1]}
        tget{t_i[2]} call{torad} cos tput{cos_ti[2]}
        tget{t_i[3]} call{torad} cos tput{cos_ti[3]}
        tget{t_i[4]} call{torad} cos tput{cos_ti[4]}
        tget{t_i[5]} call{torad} cos tput{cos_ti[5]}
        tget{t_i[6]} call{torad} cos tput{cos_ti[6]}
        tget{t_i[7]} call{torad} cos tput{cos_ti[7]}
        tget{t_i[8]} call{torad} cos tput{cos_ti[8]}
    </Macro>
```


Oren-Nayar model correction

- $A(\sigma)$ and $B(\sigma)$:

```
    </Macro>
    <Macro Name="calc_A">
        call{sigma} call{sigma} mul
        tsav(0,1)
        tget(0,1) 0.33 add
        div
        -0.5 mul
        1 add
        tput{A}
    </Macro>
    <Macro Name="calc_B">
        call{sigma} call{sigma} mul
        tsav(0,1)
        tget(0,1) 0.09 add
        div
        0.45 mul
        tput{B}
    </Macro>
```

Oren-Nayar model correction

- $\max(0, \cos(\phi_i - \phi_r))$:

```
78              <Macro Name="calc_cos_diff">
79                  call{phii} tput{phi_i}
80                  call{phir} tput{phi_r}
81                  tget{phi_i[0]} tget{phi_r[0]} sub call{torad} cos 0 max tput{cos_diff[0]}
82                  tget{phi_i[1]} tget{phi_r[1]} sub call{torad} cos 0 max tput{cos_diff[1]}
83                  tget{phi_i[2]} tget{phi_r[2]} sub call{torad} cos 0 max tput{cos_diff[2]}
84                  tget{phi_i[3]} tget{phi_r[3]} sub call{torad} cos 0 max tput{cos_diff[3]}
85                  tget{phi_i[4]} tget{phi_r[4]} sub call{torad} cos 0 max tput{cos_diff[4]}
86                  tget{phi_i[5]} tget{phi_r[5]} sub call{torad} cos 0 max tput{cos_diff[5]}
87                  tget{phi_i[6]} tget{phi_r[6]} sub call{torad} cos 0 max tput{cos_diff[6]}
88                  tget{phi_i[7]} tget{phi_r[7]} sub call{torad} cos 0 max tput{cos_diff[7]}
89                  tget{phi_i[8]} tget{phi_r[8]} sub call{torad} cos 0 max tput{cos_diff[8]}
90              </Macro>
```


Oren-Nayar model correction

- $\sin(\max(\theta_i, \theta_r))$:

```
91              <Macro Name="calc_sin_alpha">
92                  call{tr} tput{t_r}
93                  tget{t_i[0]} tget{t_r[0]} max call{torad} sin tput{sin_alpha[0]}
94                  tget{t_i[1]} tget{t_r[1]} max call{torad} sin tput{sin_alpha[1]}
95                  tget{t_i[2]} tget{t_r[2]} max call{torad} sin tput{sin_alpha[2]}
96                  tget{t_i[3]} tget{t_r[3]} max call{torad} sin tput{sin_alpha[3]}
97                  tget{t_i[4]} tget{t_r[4]} max call{torad} sin tput{sin_alpha[4]}
98                  tget{t_i[5]} tget{t_r[5]} max call{torad} sin tput{sin_alpha[5]}
99                  tget{t_i[6]} tget{t_r[6]} max call{torad} sin tput{sin_alpha[6]}
100                 tget{t_i[7]} tget{t_r[7]} max call{torad} sin tput{sin_alpha[7]}
101                 tget{t_i[8]} tget{t_r[8]} max call{torad} sin tput{sin_alpha[8]}
102             </Macro>
```

Oren-Nayar model correction

- $\tan(\min(\theta_i, \theta_r))$:

```
<Macro Name="calc_tan_beta">
    tget{t_i[0]} tget{t_r[0]} min call{torad} tan tput{tan_beta[0]}
    tget{t_i[1]} tget{t_r[1]} min call{torad} tan tput{tan_beta[1]}
    tget{t_i[2]} tget{t_r[2]} min call{torad} tan tput{tan_beta[2]}
    tget{t_i[3]} tget{t_r[3]} min call{torad} tan tput{tan_beta[3]}
    tget{t_i[4]} tget{t_r[4]} min call{torad} tan tput{tan_beta[4]}
    tget{t_i[5]} tget{t_r[5]} min call{torad} tan tput{tan_beta[5]}
    tget{t_i[6]} tget{t_r[6]} min call{torad} tan tput{tan_beta[6]}
    tget{t_i[7]} tget{t_r[7]} min call{torad} tan tput{tan_beta[7]}
    tget{t_i[8]} tget{t_r[8]} min call{torad} tan tput{tan_beta[8]}
</Macro>
```


Oren-Nayar model correction

- Main function:

```
<MainFunction>
{
in(0,3)
% Calculate cosine of illumination and save in cos_ti
call{calc_cos_ti}
% Calculate coefficient A
call{calc_A}
% Calculate coefficient B
call{calc_B}
% Calculate cosine of difference
call{calc_cos_diff}
% Calculate sin(alpha)
call{calc_sin_alpha}
% Calculate tan(beta)
call{calc_tan_beta}
% Put everything together to get the correction coefficient
tget{B} tget{cos_diff[0]} mul tget{sin_alpha[0]} mul tget{tan_beta[0]} mul tget{A} sum tget{cos_ti[0]} mul
tget{B} tget{cos_diff[1]} mul tget{sin_alpha[1]} mul tget{tan_beta[1]} mul tget{A} sum tget{cos_ti[1]} mul
tget{B} tget{cos_diff[2]} mul tget{sin_alpha[2]} mul tget{tan_beta[2]} mul tget{A} sum tget{cos_ti[2]} mul
tget{B} tget{cos_diff[3]} mul tget{sin_alpha[3]} mul tget{tan_beta[3]} mul tget{A} sum tget{cos_ti[3]} mul
tget{B} tget{cos_diff[4]} mul tget{sin_alpha[4]} mul tget{tan_beta[4]} mul tget{A} sum tget{cos_ti[4]} mul
tget{B} tget{cos_diff[5]} mul tget{sin_alpha[5]} mul tget{tan_beta[5]} mul tget{A} sum tget{cos_ti[5]} mul
tget{B} tget{cos_diff[6]} mul tget{sin_alpha[6]} mul tget{tan_beta[6]} mul tget{A} sum tget{cos_ti[6]} mul
tget{B} tget{cos_diff[7]} mul tget{sin_alpha[7]} mul tget{tan_beta[7]} mul tget{A} sum tget{cos_ti[7]} mul
tget{B} tget{cos_diff[8]} mul tget{sin_alpha[8]} mul tget{tan_beta[8]} mul tget{A} sum tget{cos_ti[8]} mul
% Calculate average correction factor
sum(9) 9 div
% Use it to correct XYZ
copy copy
div(3)
out(0,3)
}
</MainFunction>
```

Input

- The .xml file can be converted in .icc profile with iccFromXml.exe
- Applied to image with iccApplyProfiles.exe, to named input with iccApplyNamedCMM.exe
- We used:

  iccApplyNamedCMM.exe input_oren.txt 3 0 oren_nayar_correction.icc 3
- Final encoding: icEncodeFloat
- Interpolation: Linear
- Rendering intent: Absolute
- Input file:

```
1  'XYZ'    ; Data Format
2  icEncodeFloat    ; Encoding
3
4  0.5 0.5 0.5
```


Results

- Output:

```
'XYZ '   ; Data Format
icEncodeFloat    ; Encoding

;Source Data Format: 'dXYZ'
;Source Data Encoding: icEncodeFloat
;Source data is after semicolon

;Profiles applied
; oren_nayar_correction.icc

   0.8689    0.8689    0.8689   ;    0.5000    0.5000    0.5000
```

Thank you for your attention

Michele Conni
(R&D Engineer at Barbieri Electronic
Ph.D. candidate at NTNU)

michele.conni@barbierielectronic.com
michelco@stud.ntnu.no
www.barbierielectronic.com

QUESTIONS?

References

- http://www.cs.columbia.edu/CAVE/projects/btf/
- Dong, J., & Chantler, M. (2005). Capture and synthesis of 3D surface texture. International Journal of Computer Vision, 62(1-2), 177-194.
- Feidenhans, N. A., Hansen, P. E., Pilný, L., Madsen, M. H., Bissacco, G., Petersen, J. C., & Taboryski, R. (2015, October). Industrial characterization of nano-scale roughness on polished surfaces. In Optifab 2015 (Vol. 9633, p. 96330B). International Society for Optics and Photonics.
- Gonzalez, R. C., & Woods, R. E. (2002). Digital image processing [M]. Publishing house of electronics industry, 141(7).
- Haindl, M., & Filip, J. (2013). Visual texture: Accurate material appearance measurement, representation and modeling. Springer Science & Business Media.
- Luo, L., Tsang, K. M., Shen, H. L., Shao, S. J., & Xin, J. H. (2015). An investigation of how the texture surface of a fabric influences its instrumental color. Color Research & Application, 40(5), 472-482.

- Oren, M., & Nayar, S. K. (1994). Generalization of Lambert's reflectance model. In Proceedings of the 21st annual conference on Computer graphics and interactive techniques (pp. 239-246).
- Specification ICC.2:2019 (Profile version 5.0.0 - iccMAX), Image technology colour management — Extensions to architecture, profile format and data structure
- https://en.wikipedia.org/wiki/Lambertian_reflectance

ISBN 978-82-326-5381-2 (printed ver.)
ISBN 978-82-326-6733-8 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology