Towards Exploiting Change Blindness for Image Processing

Steven Le Moan, Ivar Farup, Jana Blahová

Faculty of Computer Science and Media Technology
NTNU - Norwegian University of Science and Technology
Gjøvik, Norway.

Abstract

Change blindness is a type of visual masking which affects our ability to notice changes introduced in visual stimuli (e.g. change in the colour or position of an object). In this paper, we propose to use it as a means to identify image attributes that are less important than others. We propose a model of visual awareness based on low-level saliency detection and image inpainting, which identifies textured regions within images that are the most prone to change blindness. Results from a user study demonstrate that our model can generate alternative versions of natural scenes which, while noticeably different, have the same visual quality as the original. We show an example of practical application in image compression.

Keywords: Perception, Visual Awareness, Visual Attention, Change Blindness, Saliency, Image Quality

1. Introduction

With the number of digital pictures taken every year running into the trillions [1], it has become increasingly important to understand how people perceive image contents in order to manipulate them more efficiently. Indeed, our visual system filters out visual information in a variety of ways and a wide range of image processing applications such as compression, watermarking or cross-media reproduction rely on identifying what we can and cannot see within images. For instance, very high frequency components can be removed without disturbance in typical viewing conditions due to limited contrast sensitivity, which is useful for data reduction [2]. Other early vision[1] mechanisms such as low-level texture masking [4], salience [5, 6, 7] or chromatic adaptation [8] have also been used to predict subjective image quality assessments and improve image processing techniques. On the other hand, higher levels of perception and cognition (late vision) are also subject to a number of flaws which can affect our perceptual experience and interpretation of image quality [9, 10]. While limits in our early vision renders image attributes invisible, even if we know where they are, late vision flaws pertain more to the perceived importance of these attributes. In the case of images (as opposed to videos) the distinction between *invisible* and *unimportant* is crucial in that it involves time: if a distortion cannot be detected rapidly, it can arguably be considered as acceptable. In this study, we test this hypothesis in the particular case of natural images containing large and complex textured regions. Unlike prior work on exploiting perceptual failures for prediction of image quality, we propose to identify the *important* information in images via a relatively unknown high-level mechanism of the human visual system (HVS): *visual awareness*.

[1] Typically, *early vision* refers to the first steps of visual perception where basic features like motion, colour and binocular disparity are measured [3].

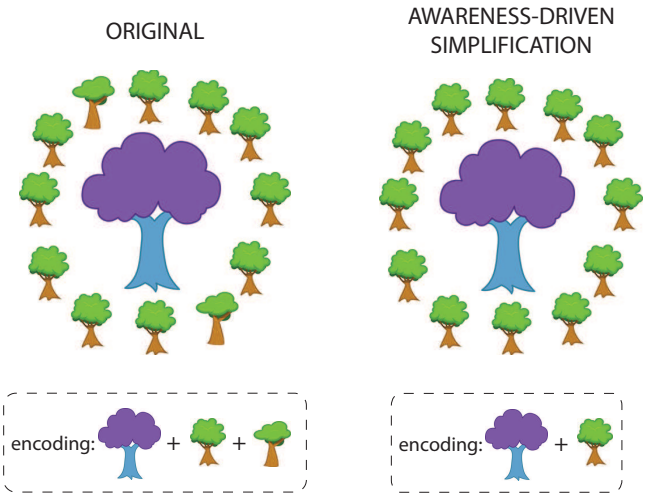


Figure 1: Principle of awareness-driven image simplification: make observers unaware that the background has been simplified by exploiting change blindness. Notice that on the left image, there are two different kinds of green/brown trees, whereas on the right, there is only one kind. As a result, the amount of data needed to encode the latter image is smaller.

Figure 2 depicts an example of the game "Spot the difference". We found that it takes most people at least 45 seconds to notice the missing engine under the wing. This comes from a remarkable shortcoming of our visual system referred to as *change blindness* [12, 13]. While the origins and implications of this phenomenon are not yet fully understood, it is known to come from a failure to accurately represent and compare visual stimuli in memory [14]. Changes affecting the gist of the scene are detected faster [15, 16], yet the nature and context of the change can make it very difficult to see, even in the most salient image regions [17, 18], as in Figure 2. Change blindness

Figure 2: Example of image pair inducing change blindness [11].

therefore highlights a difference between attention and awareness [19]. The presence of the engine in Figure 2 can then be seen as a piece of information that our visual system considers not worth verifying in priority in the context of the game. In other words, the HVS filters out the *information* that would otherwise make us rapidly aware of the change. But once the latter is noticed, it becomes clear which image is the original one and which one was tempered with, as it seems unlikely that the plane would not have an engine. Figure 3 shows another example, where change is introduced in complex textures. Again, given time, we can notice discrepancies that are invisible at first. However, this time it is less obvious which image is the original (that is of course, if one is not already familiar with the scene) as they both convey the same meaning. Consequently, we argue that the two images in Figure 3 can be considered as perceptually equivalent.

Images attributes of low-level (contrasts, edges, textures) and high-level (characteristics of objects, people, context) types are encoded in internal representations with different levels of details [20], again depending on their presumed importance in the context. If change blindness occurs, it implies that some of these attributes were not encoded in short-term memory with a sufficient fidelity [13]. Exploiting change blindness can however be a challenge. Removing the engine in Figure 2 seems difficult to do in an automatic fashion, but most importantly it has limited advantages when it comes for example to compressing the image. Other types of image attributes such as complex textures can be tempered with more easily and with greater benefits. Research on texture perception [21] and texture synthesis [22] have shown that perceptual characteristics of textures can be captured by means of only a few statistical attributes. Two different textured regions may then be perceived as similar if they match in terms of these attributes [23]. In a previous work [24], we proposed to use exemplar-based inpainting (also referred to as *image completion* or *image filling*) to simplify background textures in natural images and therefore gain in compression ratio. We demonstrated in particular that the simplified image can be considered as equivalent to the original, as long as the former is free of artifacts and semantic inconsistencies. Figure 1 illustrates the principle of exploiting change blindness to reduce the complexity of a scene's background.

Note that there is already a vast body of literature on texture masking (see e.g. [4]), yet they concern mostly the early vision aspects of texture perception. What makes our work novel is our position that it does not necessarily matter whether the original and reproduced images are noticeably different as long as it takes a while to notice the difference and that, once noticed, identifying the original one is not straightforward (as in Figure 3). We build upon previous work [24], and propose a prototypical visual awareness model to exploit spatial redundancies in textured images based on saliency detection and predictability of image regions. We first discuss related work and contributions before presenting the model as well as our experimental results.

2. Related work

2.1. Exploiting, inducing and measuring change blindness

So far, change blindness has mostly been studied in the field of vision and cognitive sciences [14, 13, 20, 25] as it received limited interest from in the image processing community [11], mostly because too little is known as to its causes thus making it a difficult phenomenon to predict/induce with in an automatic fashion. One of the very first attempts at exploiting change blindness was by Cater *et al.* [26] in the field of computer graphics. The authors suggested to lower the rendering quality of objects of lesser saliency in a scene during a visual disruption such as a blink. They conducted an experiment with 10 rendered scenes and modified the rendering quality at different locations, classified as of *central* or *marginal* interest. Central interest changes were detected rapidly while marginal interest changes required observers an average of 40s to be discovered. They concluded that change blindness can indeed be exploited in order to reduce the computational effort required for rendering.

Figure 3: Other example of image pair inducing change blindness.

When it comes to measuring the degree of change blindness between two images, Hou *et al.* [27] proposed to use the Hamming distance between two images' signatures (the signature of a greyscale image is the sign of its DCT coefficients) as a measure of the time needed by a person to actually perceive a change between the two images. Based on results obtained from an experiment involving 60 image pairs manually modified to induce change blindness and nine naive observers, they obtained an average correlation of 0.563 when the signatures were computed in the CIELAB colour-space. In addition to being only intended for large changes that significantly affect the frequency content of the image, the signature-based model is mostly *ad hoc* and lacks biological plausibility. More recently, Ma *et al.*[18] proposed a measure based on a so-called *context-aware* saliency detection and obtained a correlation between predicted degree of blindness and recognition time of 0.75 based on results obtained on 100 image pairs and 30 subjects. The changes in image pairs were generated automatically with a method that utilises a variety of operators such as insertion, deletion, replacement, scaling, etc. They used alpha matting for segmentation and PatchMatch inpainting [28] for filling when needed. However, Ma *et al.*'s model is also intended for large changes that affect objects. In this paper, we aim to induce more subtle changes which, while visible, do not significantly alter our interpretation of the scene.

2.2. Exploiting spatial redundancies in images

Several studies have suggested to exploit spatial redundancy to compress images and videos by removing macro-blocks on the encoder's side while making sure that they can be recovered on the decoder's side via inpainting-like methods [29, 30, 31]. The motivation behind these approaches is however to recover exactly the original signal, while we suggest that some discrepancies can be introduced without disturbance.

2.3. Modeling visual memory

Studies on the role of memory in the perception of visual stimuli have shown that humans have the ability to remember a massive quantity of visual information from natural scenes [32, 33]. However, not all visual information is equally remembered. In an attempt to understand what makes that some visual information will be more likely to be memorised than other, several recent studies have focused on finding the right combination of image attributes that can predict how *memorable* an image is [34] or an object in an image [35]. The results highlight in particular the importance of semantics (labels, annotations). For instance, an image containing a person or a car is more likely to be remembered than one containing a building or a tree. While these methods are primarily intended to model long-term visual memory and how internal representations fade over time, our approach is substantially different as we focus mostly on the limitations of short-term memory through the study of awareness and change blindness.

3. Contributions

As reported in [11], change blindness and other types of high-level visual masking have received limited attention from the image, video, and computer graphics research communities. Here, we test the hypothesis that change blindness can be used to identify unimportant image attributes. We build upon previous work [24], and propose a prototypical visual awareness model based on low-level feature extraction, which can be used to exploit spatial redundancies in textured images. As mentioned previously, what makes our work novel is our position that it does not necessarily matter whether the original and reproduced images are noticeably different as long as it takes a while to notice the difference and that, once noticed, identifying the original one is not straightforward.

4. Finding textured regions that can induce change blindness

4.1. Texture representation

As most natural textures can be well modeled by blocks (also referred to as patches, i.e. small neighbourhoods pixels, typi-

cally square-shaped) [36], we propose to consider the hypothesis that there is a block-based representation of textures in internal representations. This hypothesis can be supported by two demonstrated facts: 1) our vision is only detailed on a small portion of our visual field, corresponding to the size of a thumbnail seen at arm's length, which corresponds to a large block in the displayed image [23, 11] and 2) the primary visual cortex contains localised receptive fields that can be modeled by blocks [37].

4.2. Encoding fidelity map

For a digital image $\mathbf{I}$ divided in a set of non-overlapping square blocks of size n, we propose a model to estimate, for each block $\mathbf{b}$, the overall fidelity with which it will be stored in visual short-term memory noted $\mathcal{F}(\mathbf{b})$. The resulting map of all blocks in $\mathbf{I}$, noted $\mathcal{F}(\mathbf{I})$ is what we refer to as an *encoding fidelity map*. In this study, the notion of image background is particularly important. The proposed framework is indeed mostly intended for scenes that contain a (group of) prominent object(s) that constitute their foreground (e.g. the pen in Figure 3). For an image $\mathbf{I}$, we note its fore- and background $\mathbf{I_F}$ and $\mathbf{I_B}$, respectively. Incidentally, we assume that every block in $\mathbf{I_F}$ has a maximal encoding fidelity, i.e. $\mathcal{F}(\mathbf{b}) = 1, \forall \mathbf{b} \in \mathbf{I_F}$ (where 1 is the top of the fidelity scale). To extract $\mathbf{I_F}$ and $\mathbf{I_B}$, we used saliency detection and mean shift segmentation, as in [38]. This allows us in particular to reduce the computational complexity of the method by processing solely background blocks.

As previously mentioned, our model identifies those specific blocks which have both a low saliency (i.e. part of the background) and which are somehow easy for the brain to "guess" from the rest of $\mathbf{I}$. Given $\mathcal{S}(\mathbf{I})$, a saliency map derived from $\mathbf{I}$, we obtain $\mathcal{S}(\mathbf{b})$ as the average saliency of all pixels in $\mathbf{b}$. Note that the extent to which a block can be easily guessed is what we previously referred to as its *inpaintability*[2] [24]. As highlighted in previous studies [39], predicting the quality of inpainting requires to account for local context. Furthermore, it has to do so in a way that is consistent with the inpainting algorithm to be used. Let us note $\mathbf{b}_+$ the surrounds of $\mathbf{b}$ (we consider all eight 8-connected blocks of size n) and let $\mathbf{\Pi}_{\mathbf{b}_+}$ be a dictionary (set) of blocks representing $\{\mathbf{b}'_+ \in \mathbf{I} | \mathbf{b}' \notin \mathbf{I_F}\}$, the set of surrounds of all blocks not belonging to $\mathbf{I_F}$. Note that, for convenience sake, $\mathbf{\Pi}_{\mathbf{b}_+}$ is here represented as a matrix with pixel blocks (reshaped as row vectors) in rows. Finally, let ω_+ be the vector of optimal weights for the linear decomposition of $\mathbf{b}_+$ in $\mathbf{\Pi}_{\mathbf{b}_+}$, in the sense of a measure of similarity $\mathcal{D}(\mathbf{b_1}, \mathbf{b_2})$, so that:

$$\omega_+ = \arg\min_{\omega} \mathcal{D}(\mathbf{b}_+, \mathbf{\Pi}_{\mathbf{b}_+}\omega) \tag{1}$$

The encoding fidelity of block $\mathbf{b}$ is then computed as the maximum accuracy with which it can be estimated from the dictionary, weighted by its saliency:

$$\mathcal{F}(\mathbf{b}) = \mathcal{D}(\mathbf{b}, \mathbf{\Pi}_{\mathbf{b}}\omega_+)\mathcal{S}(\mathbf{b}) \tag{2}$$

[2] We define the *inpaintability* of a set of pixels as the probability that it can be replaced in a visually appealing manner, with a given inpainting method.

where $\mathbf{\Pi}_{\mathbf{b}}$ is a dictionary of blocks corresponding to $\mathbf{\Pi}_{\mathbf{b}_+}$, but from the set $\{\mathbf{b}' \in \mathbf{I} \setminus \mathbf{I_F}\}$ (i.e. without the surrounds). Figure 4 shows an example of resulting map.

Many different methods have been proposed to create a dictionary of patches or blocks of pixels for image denoising, restoration or inpainting [40]. Any of them can potentially be used in our model. In this study however, and for the sake of simplicity we chose to use Principal Component Analysis (PCA) to build $\mathbf{\Pi}_{\mathbf{b}}$ and $\mathbf{\Pi}_{\mathbf{b}_+}$. The first 50 principal components were kept in both cases. To represent colour, we use the hue-linearised LAB2000HL colour space [41], which exhibits more perceptual uniformity than CIELAB overall.

4.3. Saliency detection

There is a vast literature on saliency detection for visual attention modeling [6]. Although visual attention is known to be a relatively more complex process that involves also top-down mechanism such as culture or personal preference, bottom-up saliency models have been reported to give very accurate prediction of human fixations in some cases. We tested several models [42, 43, 38, 44] and found that the seminal Itti model [42] with basic features (colour opposition, luminance and orientations) gives satisfying results on our experimental benchmark.

4.4. Block similarity measure

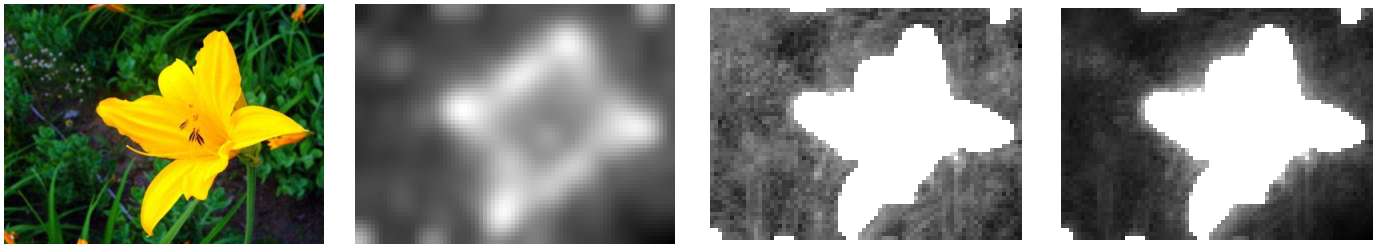
As a measure of block similarity $\mathcal{D}(\mathbf{b_1}, \mathbf{b_2})$, it is common to use the sum of squared differences [45]. Bugeau *et al.*[46] observed however that the SSD, when used alone, tends to favor uniform blocks, therefore we propose to use it in combination with a contrast and a structure similarity terms derived from the well-known SSIM index [47] such as:

$$\mathcal{D}(\mathbf{b_1}, \mathbf{b_2}) = \frac{SSD(\mathbf{b_1}, \mathbf{b_2})}{c(\mathbf{b_1}, \mathbf{b_2})s(\mathbf{b_1}, \mathbf{b_2})} \tag{3}$$

where $c(\mathbf{b_1}, \mathbf{b_2})$ and $s(\mathbf{b_1}, \mathbf{b_2})$ are respectively the contrast and structural similarity terms.

5. Modifying the least significant image regions

Having identified blocks of lower significance, we can alter them in a way that induces change blindness. Let us consider a threshold τ of internal encoding fidelity, so that the set $\mathbf{\Lambda}_\tau = \{\mathbf{b} \in \mathbf{I} | \mathcal{F}(\mathbf{b}) < \tau\}$ represents the least significant blocks in $\mathbf{I}$. We can then discard $\mathbf{\Lambda}_\tau$ and recover it via inpainting. We implemented a simple method based on the seminal work by Criminisi *et al.*[48]. The image with missing regions is first divided into blocks of size n, as in the encoding fidelity map. Every block $\mathbf{b}$ is then examined together with its eight neighbors (the surround $\mathbf{b}_+$), thus creating what we will refer to as a *super-block* $\mathbf{B} = \{\mathbf{b} \cup \mathbf{b}_+\}$. If $\mathbf{b}$, the center of $\mathbf{B}$ is missing, the super-block is put in a dictionary of incomplete blocks $\mathbf{\Pi}_\circ$. In the alternative case, the super-block is put in a dictionary of complete blocks $\mathbf{\Pi}_\bullet$. For each incomplete element in $\mathbf{\Pi}_\circ$, the partial data is used for context matching with $\mathbf{\Pi}_\bullet$ and missing data is copied from the best matching complete super-block. In



(a) (b) (c) (d)

Figure 4: From left to right: original image (a), saliency map according to [42] (high energy = high saliency) (b), encoding fidelity map without (c) and with (d) saliency weighting (high energy = high encoding fidelity). The encoding fidelity of every block in the scene's foreground (the yellow flower), which is extracted in the first steps of the map computation, is maximal.

the inpainting procedure, priority is given to the most complete super-blocks. To give a sense of continuity between the known and filled regions, we quilted blocks by means of graphcut [49] and Poisson blending [50].

The higher the threshold τ, the lower the chance to recover exactly $\mathbf{\Lambda}_\tau$, as more blocks need to be recovered from less reference data. However, as long as there are no artifacts or semantic inconsistencies, we argue that the result is acceptable. Unfortunately, traditional image-difference metrics[3] do not predict well such inconsistencies [24], therefore we still have to rely on a manual selection of τ. However, our results demonstrate that, for a certain type of images, a threshold of 10% permits to render images with a similar quality or higher to that of the original, according to a majority of people.

6. Experiments

6.1. Viewing Conditions

We used an Eizo colourEdge CG246W display (24.1" - 61cm), calibrated with an EyeOne software for a colour temperature of 6500K, a gamma of 2.2 and a luminous intensity of 80cd/m^2. The experiment was carried out in a dark room. A viewing distance of approximately 50cm was ensured for all observers.

6.2. Observers

A group of 30 colour-normal observers participated in the experiment. Ages ranged from 22 to 52 years old, 20 of them were male and 15 of them had background in image processing or vision research. Note that we found no significant correlation between the output of the experiment and either of these criteria (age, gender and familiarity with the task).

6.3. Stimuli and Methodology

In order to assess the extent to which the proposed model can induce change blindness, we used 30 colour images of natural scenes consisting of a complex textured background with a spatially compact foreground (see Figure 5). These images were selected from two publicly available databases [51, 52], except one which was computationally rendered specifically for this study in order to show that the proposed method also works with non-natural images. For each scene, three modified versions of the original image were rendered for $\tau = 5\%$, $\tau = 10\%$ and $\tau = 15\%$. Initially, the image pairs (original, rendering at $\tau = 5\%$) were displayed in random order and position (left/right) and observers were asked to answer the question: *Which image has the highest quality?*, with the possibility of tie scores. Participants were given no indication as to the meaning of the term *quality*, it was left entirely up to them to interpret it. We expected people to occasionally see some differences between the stimuli after a while, when the blindness would stop. Therefore, we could not ask them to rate the *fidelity* or *difference* for instance. As long as they were not able to tell which was the original which was simplified, the framework was successful. In case a participant could see inpainting artifacts, the original image was most likely considered by them as reference and the other, distorted. These are the reasons why we chose to use the word *quality*.

Furthermore, for each scene and at each level, a pair consisting of twice the original image (i.e. $\tau = 0\%$) was also introduced at a random position in the experiment. The purpose of these red herrings was to test our initial intuition that, in the absence of difference between the stimuli, observers could somehow convince themselves that they saw a difference, partly due to the same short-term visual memory flaws that induce change blindness. Tie scores were also allowed in these cases. We then also counted the number of occurrences of each possible outcome across all observers and image pairs: $\texttt{count}_4$ for each time one of the two identical images was found of higher quality than the other and $\texttt{count}_5$ for when a tie score was given.

Note that we constrained the sequence order to show scenes in order of increasing levels of simplification (i.e. $\tau = 5\%$ then 10% and finally 15%), in order to avoid the perception of artifacts at the highest level influencing judgment at lower levels. A screening according to [53] revealed that all observers were valid, which implies that there was some consistency between their judgments.

Unlike in our preliminary work [24], we focus on demonstrating that observers find no difference in terms of *quality* between original and simplified images. This approach allows us

[3] Note that the term *metric* is here not used according to its proper mathematical definition. It is however quite established in the image quality community.

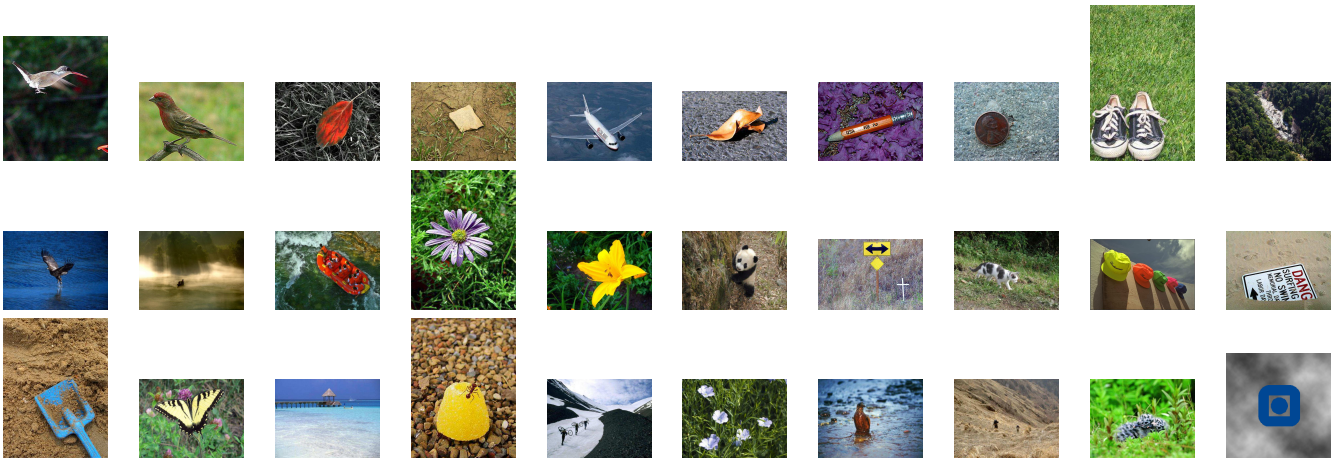


Figure 5: Images used in our experiments, ranked from the one yielding the best (top left) to the worst (bottom right) quality of simplification at $\tau = 15\%$ according to our users panel. These images were selected from two publicly available databases [51, 52], except the last one, which was rendered specifically for this study.

to demonstrate that, even if observers can see a discrepancy between the two stimuli, the simplified one can be perceived as of equivalent image quality to the original.

6.4. Results

Table 1 gives the results obtained for all 30 images and 30 observers. In order to demonstrate the efficiency of our method, we compared the probability of an observer finding the original image to be of higher quality to that of the same observer finding the simplified image to be of equivalent or higher quality than the original one. First, as a model of standard observer, we simply looked at each image pair and computed the mode of the decision taken by all observers (i.e. the majority). The two probabilities can then be estimated from our experimental data by the ratios of the number of occurrences of each case (i_1 and i_2) over the total number of comparisons $m = 120$ (4 values of τ - 0, 5, 10 and 15% - $\times$ 30 scenes $\times$ one standard observer), i.e.: $p_1 \approx \hat{p}_1 = i_1/m$ and $p_2 \approx \hat{p}_2 = i_2/m$, respectively.

To determine whether these estimated probabilities were *significantly* different from each other, we assumed that observers' ability to find the original in each of the image pairs follows a binomial distribution and used Yule's two-sample binomial test [54] at 95% confidence.

The test revealed that, when $\tau = 5\%$ and $\tau = 10\%$, observers did not find that the original image was of higher quality. However, when $\tau = 15\%$, they did. Additionally, our results indicate that, when asked to compare two identical images, the probability that observers found differences in terms of quality between them is not significantly different from that of not seeing any difference. This is particularly interesting as it challenges the mainstream approach to subjective quality assessment. Though recent studies have stressed the importance of considering the multiple strategies employed by our visual system when assessing the resemblance of an image pair [55], our results reveal that people can as well *hallucinate* the presence of image distortions.

Figure 6 shows examples of best and worst results obtained, as evaluated by our panel of observers.

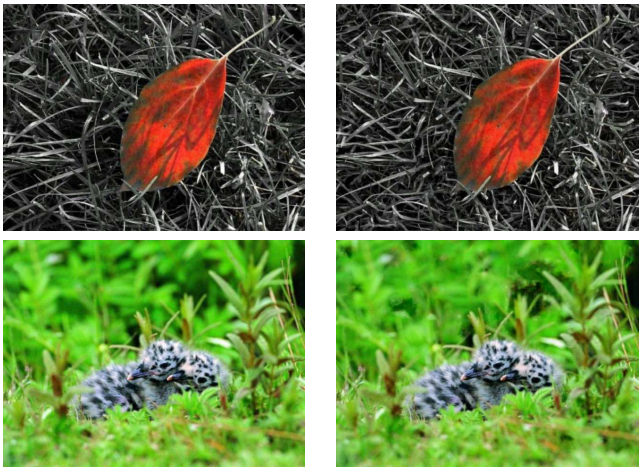


Figure 6: Example of good (first row) and bad (second row) results obtained at $\tau = 15\%$, according to the observers ratings. Note that, despite using a small block size (n=8) to describe the textures, the coarse background of the top scene could still be simplified with excellent quality.

6.5. Time analysis

In terms of decision times, we measured relatively large inter-observer variability. The whole experiment took between 6.3 and 58.4 minutes with an average at 23.4 minutes (i.e. respectively 3.13, 29.2 and 11.7 seconds per image pair). We also measured a large inter-scene variability with an average standard deviation per observer of 5.2s at $\tau = 15\%$ (and statistically similar values at $\tau = 5\%$ and $\tau = 10\%$[4]). We believe that this can be explained by the fact that the scenes differ greatly, not

[4]Two mean values m_1 and m_2 with corresponding standard deviations σ_1 and σ_2 are considered to be significantly different if $m_1 + \sigma_1 < m_2 - \sigma_2$ (if $m_1 < m_2$) or $m_2 + \sigma_2 < m_1 - \sigma_1$ (if $m_1 > m_2$).

Table 1: Results from the subjective experiments in percentage of total number of image pairs. **O**: original image, **S**: simplified image. Note that (1), (2) and (3) add up to 100% in each column. Values in bold are significantly larger than their counterpart (e.g. (1) is the counterpart of (2)+(3)). These results show that 1) the simplified images were considered at least as good as the original ones quality-wise in a majority of the cases for τ =5% and τ =10% and 2) differences between image pairs were occasionally hallucinated.

	τ =0% (*red herring*)	τ =5%	τ =10%	τ =15%
(1) **O** was preferred	27%	29%	45%	**53%**
(2) **S** was preferred	24%	26%	23%	21%
(3) No difference	49%	45%	32%	26%
(1)+(2): Difference was hallucinated	51%	/	/	/
(2)+(3): **S** was considered of equal or better quality than **O**	/	**71%**	**55%**	47%

only in terms of the textures they contain (colour, coarseness, orientation, etc) but also in terms of size, location compactness and meaning of their foreground. Not only can each of these attributes affect the performance of our method, they can also significantly affect the time needed to perform the subjective task.

Figure 7 depicts the image pairs that required the most and least time to rate, on average. Note finally that we found no Pearson Correlation Coefficient larger than 0.55 between decision types and decision times (globally or per observer), meaning that these variables correlate poorly.

Figure 7: Image pairs that required the most (top) and least (bottom) time for a decision, on average for 30 observers (original images are shown on the left). The levels of simplifications are: 5% (top) and 15% (bottom). The average times recorded are 16.7s (σ = 11.7s) and 5.2s (σ = 3.4s).

7. Discussion

The results presented in the previous section demonstrate that we can "simplify" complex natural textures by carving out up to 10% of data without perceivable loss of quality, in certain types of scenes.

In broad terms, change blindness can be exploited with any scene containing more *information* than one can store in visual short-term memory. In our framework, *information* is defined as details within rich textures and the quality of its results depends mostly on a trade-off between three attributes of textured regions: richness, stationarity and size. The richer the texture, the more unlikely people are to be aware of all its details. Consider the simplest possible case of an image with all its pixels of the exact same colour: there is then no *information* to be missed, so no perceptual failure to exploit. On the other hand, the texture needs to be sufficiently stationary so it can be synthesised from a small number of representative patches. The larger the textured region, the larger the number of blocks to be potentially removed (and the better the compression ratio). Our framework considers 8x8 pixel blocks and 24x24 pixel superblocks so it should be applied to scenes with at least one textured region larger than a 24x24 pixel square.

The optimal value of τ is determined by the richness, stationarity and size of textured regions within the scene. For instance, when some patterns seem to be duplicated at different locations with small variations (refer to Figure 1 for an illustration). The larger, richer and more stationary the textures are, the larger the optimal τ. If poor quality results are obtained with τ=5% for a particular image, it is likely that the framework is not suitable for it.

Of course, the proposed framework relies heavily on the performance of the texture synthesis and saliency detection. Other inpainting strategies may be more adapted for other kinds of scenes [45] and/or may allow for higher degrees of simplification. We believe that this constitutes the most relevant direction for future research to improve this work. Similarly, the use of other saliency detection models (see [6]) may be advisable depending on the type of image under consideration.

In addition to providing insights into visual coding, these findings can be used in block-based compression such as with JPEG or HEVC. The fewer the blocks, the smaller the amount of data to encode and consequently the better the compression ratio. We analysed empirically how our framework can improve the performance of JPEG and HEVC coding (with the Main Still Picture profile for the latter) on the benchmark set from Figure 5 as well as the Kodak Lossless True colour Image Suite [56], with a block/coding tree unit size of 8x8 pixels. The two coding schemes were implemented in Matlab and their re-

spective compression ratios were tuned manually so as to create high quality reproductions (we measured an SSIM index value larger than 0.96 in each case). The framework was then applied to each image in order to reduce the number of blocks to encode, resulting in an improvement in compression ratio. Our results, reported in Table 2, indicate a marginal yet significant average improvement over both standard JPEG and HEVC. Of course,

Table 2: Average improvements of compression ratios permitted by our framework on two different datasets. The framework was applied to JPEG and HEVC (Main Still Picture profile with coding tree units of size 8x8). Recall that, as per our experimental results, the framework can produce high quality images for $\tau \leq 10\%$.

		τ (%)		
		5	10	15
Benchmark	JPEG	+3.7%	+6.9%	+10.6%
	HEVC	+3.3%	+5.2%	+9.0%
Kodak LTCIS	JPEG	+3.1%	+6.2%	+9.8%
	HEVC	+2.1%	+4.7%	+8.1%

The change blindness-inducing framework that we propose can also be exploited for fragile *watermarking* and other security applications. Indeed our experimental results demonstrate that, in a significant number of cases, people were not able to distinguish between the original and reproduced images. Therefore, we can create several unique versions of the same image that no one would even perceive as different, at least not at first glance. Instead of removing as many blocks as possible as for compression, we can select a unique pattern/combination of blocks to be simplified, and this will basically constitute the watermark. The whole point is that the location of these blocks would always be unknown to the receiver, who would then find it very challenging to fraudulently alter the watermark other than by altering the entire file. This is of course applicable to videos as well.

Finally, we also showed that people can hallucinate[5] discrepancies in terms of quality between two identical natural images. Though it is not completely clear which perceptual/psychological mechanisms are behind this remarkable phenomenon, we believe that it challenges the mainstream approach to subjective image quality assessment (SIQA) in that it reveals a type of subjective bias which has not yet been accounted for in image quality models, especially when it comes to near-threshold distortions [55].

8. Conclusions

We demonstrated how visual change blindness can affect subjective tasks pertaining to image quality assessment and we proposed a bottom-up model of visual awareness in order to predict it. Results from a user study revealed that we can alter up to 10% of pixels within certain types of images without perceivable loss of quality. We then demonstrated that this is exploitable for image coding as it can improve compression ratios for block-based approaches like JPEG and HEVC. Our findings call for more investigations towards understanding change blindness, visual awareness and how it affects what we see in natural images or video sequences.

[5]Here we use the word "hallucinate" in a broad sense, not implying specifically that the failure occurs at a purely perceptual level.

9. Acknowledgment

This research was funded by the Research Council of Norway (SHP project 221073).

References

[1] C. Cakebread, People will take 1.2 trillion digital photos this year thanks to smartphones, Business Insider.
[2] A. Skodras, C. Christopoulos, T. Ebrahimi, The JPEG 2000 still image compression standard, Signal Processing Magazine, IEEE 18 (5) (2001) 36–58.
[3] E. Adelson, J. Bergen, et al., The plenoptic function and the elements of early vision.
[4] M. Alam, K. Vilankar, D. Field, D. Chandler, Local masking in natural images: A database and analysis, Journal of vision 14 (8) (2014) 22–22.
[5] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2010) 185–198.
[6] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. on Pattern Analysis and Machine Intelligence 35 (1) (2013) 185–207.
[7] L. Zhang, Y. Shen, H. Li, VSI: A visual saliency induced index for perceptual image quality assessment, IEEE Trans. Image Process. 23 (10) (2014) 4270–4281.
[8] J. Preiss, F. Fernandes, P. Urban, Color-image quality assessment: From prediction to optimization, IEEE Transactions on Image Processing 23 (3) (2014) 1366–1378.
[9] S. Le Moan, M. Pedersen, I. Farup, J. Blahová, The influence of short-term memory in subjective image quality assessment, in: International Conference on Image Processing, IEEE, 2016, pp. 91–95.
[10] S. Le Moan, M. Pedersen, Evidence of change blindness in subjective image fidelity assessment, in: International Conference on Image Processing, IEEE, 2017.
[11] C. Healey, J. Enns, Attention and visual memory in visualization and computer graphics, IEEE Trans. Visual Comput. Graphics 18 (7) (2012) 1170–1188.
[12] D. J. Simons, M. S. Ambinder, Change blindness theory and consequences, Current directions in psychological science 14 (1) (2005) 44–48.
[13] M. Jensen, R. Yao, W. Street, D. Simons, Change blindness and inattentional blindness, Wiley Interdiscip. Rev. Cognit. Sci. 2 (5) (2011) 529–546.
[14] D. A. Varakin, D. T. Levin, K. M. Collins, Comparison and representation failures both cause real-world change blindness, Perception 36 (5) (2007) 737–749.
[15] R. A. Rensink, J. K. O'Regan, J. J. Clark, To see or not to see: The need for attention to perceive changes in scenes, Psychological science 8 (5) (1997) 368–373.
[16] J. K. O'Regan, R. A. Rensink, J. J. Clark, Change-blindness as a result of 'mudsplashes', Nature 398 (6722) (1999) 34–34.
[17] J. A. Stirk, G. Underwood, Low-level visual saliency does not predict change detection in natural scenes, J. Vis. 7 (10) (2007) 3.
[18] L.-Q. Ma, K. Xu, T.-T. Wong, B.-Y. Jiang, S.-M. Hu, Change blindness images, Visualization and Computer Graphics, IEEE Trans. on 19 (11) (2013) 1808–1819.
[19] V. Lamme, Why visual attention and awareness are different, Trends in cognitive sciences 7 (1) (2003) 12–18.
[20] T. F. Brady, T. Konkle, G. A. Alvarez, A review of visual memory capacity: Beyond individual items and toward structured representations, J. Vis. 11 (4) (2011) 1–34.

[21] B. Julesz, Textons, the elements of texture perception, and their interactions, Nature 290 (5802) (1981) 91–97.
[22] G. Tartavel, Y. Gousseau, G. Peyré, Variational texture synthesis with sparsity and spectrum constraints, J. Math. Imaging Vision 52 (1) (2015) 124–144.
[23] J. Freeman, E. Simoncelli, Metamers of the ventral stream, Nat. Neurosci. 14 (9) (2011) 1195–1201.
[24] S. Le Moan, I. Farup, Exploiting change blindness for image compression, in: 11th International Conference on Signal, Image, Technology and Internet Based Systems (SITIS), IEEE, Bangkok, Thailand, 2015, pp. 1–7.
[25] M. A. Cohen, D. C. Dennett, N. Kanwisher, What is the bandwidth of perceptual experience?, Trends in Cognitive Sciences 20 (5) (2016) 324–335.
[26] K. Cater, A. Chalmers, C. Dalton, Varying rendering fidelity by exploiting human change blindness, in: Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia, ACM, Melbourne, Australia, 2003, pp. 39–46.
[27] X. Hou, J. Harel, C. Koch, Image signature: Highlighting sparse salient regions, Pattern Analysis and Machine Intelligence, IEEE Trans. on 34 (1) (2012) 194–201.
[28] C. Barnes, E. Shechtman, A. Finkelstein, D. Goldman, Patchmatch: A randomized correspondence algorithm for structural image editing, ACM Trans. on Graphics-TOG 28 (3) (2009) 24.
[29] S. D. Rane, G. Sapiro, M. Bertalmio, Structure and texture filling-in of missing image blocks in wireless transmission and compression applications, IEEE Trans. on Image Processing 12 (3) (2003) 296–303.
[30] Z. Xiong, X. Sun, F. Wu, Block-based image compression with parameter-assistant inpainting, IEEE Trans. on Image Processing 19 (6) (2010) 1651–1657.
[31] F. Racapé, O. Déforges, M. Babel, D. Thoreau, Spatiotemporal texture synthesis and region-based motion compensation for video compression, Signal Process. Image Commun. 28 (9) (2013) 993–1005.
[32] L. Standing, Learning 10000 pictures, The Quarterly journal of experimental psychology 25 (2) (1973) 207–222.
[33] T. Brady, T. Konkle, G. Alvarez, A. Oliva, Visual long-term memory has a massive storage capacity for object details, Proceedings of the National Academy of Sciences 105 (38) (2008) 14325–14329.
[34] P. Isola, J. Xiao, A. Torralba, A. Oliva, What makes an image memorable?, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 145–152.
[35] A. Khosla, J. Xiao, A. Torralba, A. Oliva, Memorability of image regions, in: Advances in Neural Information Processing Systems, 2012, pp. 305–313.
[36] A. A. Efros, T. K. Leung, Texture synthesis by non-parametric sampling, in: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, Vol. 2, IEEE, 1999, pp. 1033–1038.
[37] B. Olshausen, et al., Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607–609.
[38] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2009, pp. 1597–1604.
[39] J. Kopf, W. Kienzle, S. Drucker, S. B. Kang, Quality prediction for image completion, ACM Trans. on Graphics (TOG) 31 (6) (2012) 131.
[40] M. J. Gangeh, A. K. Farahat, A. Ghodsi, M. S. Kamel, Supervised dictionary learning and sparse representation-a review, arXiv preprint arXiv:1502.05928.
[41] I. Lissner, P. Urban, Toward a unified color space for perception-based image processing, IEEE Trans. on Image Processing 21 (3) (2012) 1153–1168.
[42] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.
[43] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Advances in neural information processing systems 19 (2007) 545.
[44] L. Zhang, Z. Gu, H. Li, SDSP: A novel saliency detection method by combining simple priors., in: ICIP, Citeseer, 2013, pp. 171–175.
[45] C. Guillemot, O. Le Meur, Image inpainting: Overview and recent advances, Signal Processing Magazine, IEEE 31 (1) (2014) 127–144.
[46] A. Bugeau, M. Bertalmío, V. Caselles, G. Sapiro, A comprehensive framework for image inpainting, IEEE Trans. on Image Processing 19 (10) (2010) 2634–2645.
[47] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: From error measurement to structural similarity, IEEE Trans. Image Processing 13 (4) (2004) 600–612.
[48] A. Criminisi, P. Pérez, K. Toyama, Region filling and object removal by exemplar-based image inpainting, IEEE Trans. on Image Processing 13 (9) (2004) 1200–1212.
[49] V. Kwatra, A. Schödl, I. Essa, G. Turk, A. Bobick, Graphcut textures: image and video synthesis using graph cuts, in: ACM Trans. on Graphics (ToG), Vol. 22, ACM, 2003, pp. 277–286.
[50] P. Pérez, M. Gangnet, A. Blake, Poisson image editing, in: ACM Trans. on Graphics (TOG), Vol. 22, ACM, 2003, pp. 313–318.
[51] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, IEEE Trans. on Pattern Analysis and Machine Intelligence 33 (2) (2011) 353–367.
[52] J. Li, M. D. Levine, X. An, X. Xu, H. He, Visual saliency based on scale-space analysis in the frequency domain, Pattern Analysis and Machine Intelligence, IEEE Trans. on 35 (4) (2013) 996–1010.
[53] I.-R. BT.500-12, Recommendation: Methodology for the subjective assessment of the quality of television pictures, 1993.
[54] L. Brown, X. Li, Confidence intervals for two sample binomial distribution, Journal of statistical planning and inference 130 (1-2) (2005) 359–375.
[55] E. Larson, D. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 011006. doi:10.1117/1.3267105.
URL `http://link.aip.org/link/?JEI/19/011006/1`
[56] Kodak lossless true color image suite: `http://r0k.us/graphics/kodak/` (accessed on 25/11/2017).