Importance of region-of-interest on image difference metrics

Marius Pedersen

Master's Thesis
Master of Science in Media Technology
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2007

Avdeling for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Faculty of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Abstract

Many image difference metrics have been developed in the last 4 decades. All of these metrics are constructed to predict perceived image difference, but none have been successful. When we rate image difference we look at different areas in the image, based on the difference in these areas we make a decision of the perceived difference. Information about what draws attention and how we examine images can be used to improve image difference metrics.

This research project investigates the importance of region-of-interest on image difference metrics. Region-of-interest has been extracted by using an eye tracker, but also by manual marking by the observers. 3 different tasks were performed by the observers while their gaze position was recorded. Further a manual marking of region-of-interest together with a questionnaire to map background knowledge was carried out. The information found on how we perceive and examine images has been applied to different image difference metrics, such as ΔE^*_{ab}, S-CIELAB, iCAM, SSIM and the hue angle algorithm. The issues regarding how observers look at images given different tasks are also discussed and analyzed.

The results indicate that region-of-interest improves image difference metrics, especially when the metrics already have a low performance in term of linear correlation between perceived and calculated difference. There are no clear evident that one type of region-of-interest outperform other types. The improvement in performance is therefore both scene and metric dependent.

Results also show that observers have different areas of attention according the task given to them, as freeview, rating image difference and marking important regions. The common denominator within every task is faces, and this is clearly important in all tasks for the observers. Within areas of attention will change whether the observer is an expert or non-expert.

Sammendrag

Mange bildeforskjellsmetrikker har sett dagens lys de siste tiår. Alle disse metrikkene har som mål å prediktere oppfattet bildeforskjell, men ingen har vært suksessfulle. Når mennesker klassiferer bildeforskjeller ser vi på forskjellige områder, og vi gjør oss opp en mening av den oppfattede forskjellen basert på disse områdene. Informasjonen om hva som tiltrekker seg synet og hvordan man gransker bilder skaffer oss viktig informasjon om bildeforskjeller.

Dette forskningsprosjektet undersøker viktigheten av interesseområder i bildeforskjellsmetrikker. Interesseområder er funnet ved hjelp av en eye tracker, men også ved at observatøren marker interesseområder manuelt. 3 forskjellige oppgaver ble gjennomført av observatørene samtidig som deres blikk posisjon ble registrert, i tillegg markerte de interesseområder manuelt sammen med ett spørreskjema. Informasjonen om hvordan vi oppfatter og gransker bilder har blitt påført forskjellige bildeforskjellsmetrikker, som ΔE^*_{ab}, S-CIELAB, iCAM, SSIM og en fargetone algoritme. Området som tar for seg hvordan observatører ser på bilder gitt forskjellige oppgaver blir også diskutert og analysert.

Resultatene indikerer at interesseområder forbedrer bildeforskjellsmetrikker, spesielt de metrikker som prestrer dårlig fra før med tanke på lineær korrelasjon mellom oppfattet og prediktert forskjell. Det er ingen klare bevis for at en type av interesseområder er bedre enn andre. Forbedringen er derfor både scene og metrikk avhengig.

Resultatene viser også at observatører har forskellige interesseområder for forskjellige oppgaver, som fri observasjon, vurdere bildeforskjell og markere viktige områder. Fellesnevneren i alle oppgaver er ansikter, og dette er en viktig faktor for observatørene. Innenfor hver oppgave så forandres interesseområde avhengig av om observatøren er ekspert eller novise.

Preface

This master thesis has been carried out together with The Norwegian Color Research Laboratory and the Department of Computer Science and Media Technology at Gjøvik University College. The main goal of this thesis is to investigate the importance of region-of-interest on image difference metrics.

During the work of this thesis I have gotten many good advice from my fellow students and employees at Gjøvik University College. First of all I would like to thank my supervisors, professor Jon Yngve Hardeberg and assistant professor Peter Nussbaum, for excellent supervision, many good advice, motivation and solutions. Thanks to my supervisors for encouraging me to submit parts of this work for publications.

I would also like to thank everyone who participated in my experiment, have you all been a crucial piece of this thesis. Thanks to Trond Viggo Bjerke for the creation of one of the images used.

Marius Pedersen, 2007/06/30

Contents

List of Figures

List of Tables

1 Introduction

1.1 Introduction

This section covers an introduction and background for the thesis. It also presents the research questions and justification for carrying out this thesis.

1.2 Topics covered

The main goal of this thesis is to investigate the effect of region-of-interest (ROI) on image difference metrics. Different ROIs will be used and the difference between these will be investigated. The criteria that observers build their evaluation on will also be analyzed.

1.3 Background

The quality of an image and the difference between an original and a reproduction is something that concerns people in different situations, both private users and in the industry. When we print an image we want the output to be as close to the original as possible. Are changes made to an image perceivable for the observers, and to what degree? Image difference metrics have been developed to answer this question, their goal is to predict the perceived image difference.

The image difference metrics used today do not predict the perceived image difference very well. The attempt to develop a metric working under different conditions have not been successful. The complexity of the human visual system cannot be simulated well enough to predict the perceived image difference.

The importance of eye movements in visual perception has been recognized in several studies. By using an eye tracker the visual paths and fixations point of an observers can be found. Important regions and how they attract fixations can be used as useful information when predicting image difference.

1.3.1 Problem description

Today there is no good way of predicting perceived image quality without doing a psychophysical experiment. This is both time consuming and resource demanding. Image difference metrics available today have been proved shortcoming when it comes to predicting perceived image difference. In an image some regions are more important than others. These regions are both observer and scene dependent. The regions that are important and draws most attention should be taken into account and weighted higher in a pixel-by-pixel image difference metric.

If important regions could be identified, changes in an image could be done according to the importance of the regions, and this could improve the overall perceived image quality or image difference.

The aim of this project is to investigate the importance of region-of-interest on image difference metrics. Comparison of different regions from the eye-tracker and observers are also carried out.

1.4 Justification, motivation and benefits

Many people, both in the graphic art business and the average user, would like to predict image difference, and how perceivable a modification is to the observer. This would eliminate the use of time consuming and resource demanding psychophysical experiments. Stakeholders for this thesis will be anyone interested in predicting image quality or image difference. This could range from companies to average users wanting to know how much changes in an image are perceived by the observer. Also how people look at images and what regions that are important to them will be important to photographers, the graphic art business and the average user.

When a change is made in an image, which regions of the images do the observers pay attention to? Could a great color difference in some regions be acceptable and just noticeable, while in other regions small color difference will be well perceivable? Knowing this in beforehand can help us to predict perceived image difference.

By identifying region-of-interest and its importance to image difference metrics can introduce new ways of developing image difference metrics. The results will answer if incorperating ROI algorithms to the framework of image difference metrics will be useful. Research has been done in the field of automatically region-of-interest simulating human region-of-interest, and if an improvement is found by using eye tracking further research can be done in this field.

1.5 Research questions

2 research questions have been formulated as the basis for this thesis.

- Q1: Can region-of-interest improve overall image difference metrics in complex images?
- Q2: How do observers look at images given different tasks?

1.5.1 Q1: Can region-of-interest improve overall image difference metrics in complex images?

Today's image difference metrics do not predict perceived image difference well. Observers look at different regions in an image to base their decision upon. Can these regions be used to improve todays existing image difference metrics? To answer this region-of-interest must be identified, and then applied to different image difference metrics. The results are compared to the image difference from the metrics of the total image without the regions in connection with the results from a psychophysical experiment to decide if there is an improvement. If an improvement is found this could open the way for region-of-interest algorithms simulating human gazing in the framework of image difference metrics.

1.5.2 Q2: How do observers look at images given different tasks?

In addition to applying region-of-interest to image difference metrics, the regions alone are important and provide useful information about how we look at images. Different tasks performed by the observers can results in different important regions. Is there a difference in what we mark and look at between these tasks? Can some areas be defined as more important than others? An investigation of how we look at images given different tasks is performed.

1.6 Thesis structure

First an introduction to the state of the art in this field is given, going through the most important aspects related to this thesis. Then an overview of the experimental setup will be presented, both before and during the experiment. A chapter on the experimental results is given, followed by two main sections of analysis and results, ”How do we look at images” and ”Image difference metrics”. In Chapter 7 a conclusion is given followed by further research opportunities and a bibliography. The last section is an appendix containing supplementary data for this thesis.

2 State of the art

This chapter gives an overview for the research relevant to this thesis, and how far the research has come in the field of region-of-interest and image difference metrics.

2.1 How do we look at images

Henderson et al. [1] found that observers will acquire object properties within a radius of 4° of the fixation point. This was done by showing 33 realistic three-dimensional rendered color scenes several times with different objects removed from the scene. They also found a tendency for observers to sooner fixate on a changed region than for unchanged regions.

Jaimes et al. [2] did research on eye movements in automatic image classifiers. The results here indicate that observers have similar viewing patterns for different images in same semantic category. They also found that viewing patterns across observers in the same image were similar, but idiosyncratic behavior was present. In more complex scenes (lanscape, crowded scenes etc) there was little consistency, but in less complex scenes the viewing pattern consistency was higher.

Underwood and Foulsham [3] found that highly salient objects attracted fixations earlier than less conspicuous objects. They also discovered that when asking the observer to find an object, the visual saliency of a non-target object did not influence fixations.

Rajashekar et al. [4] used a combination of eye tracking and principal component analysis to extract low-level image features that attract human fixations in a scene. The results from this experiment indicate that humans are not random in their decision where to fixate. The same is also found by Buswell [5], where observers fixated at the same locations but not necessary in the same temporal order.

Rajashekar et al. [6] also found that observers are not random in their search task but uses a different set of strategies. This was done on simple geometric targets with different amount of noise.

Endo et al. [7] showed that individual distribution of gazing points were very similar among observers in the same scenes. The results also indicate that each image has a particular gazing area, particulary images containing human faces. They also concluded that image quality is influenced by the quality of the gazing area.

Mackworth and Morandi [8] found that a few regions in the image dominated the data. Informative areas had a tendency to recieve clusters of fixations. Half to two-thirds of the image receive few or non fixations, these areas (for example texture) was predictable, usual and not very informative.

Parker [9] found that observers tended to look at changed objects earlier than unchanged objects. Findings also support the conclusion of observers using information in the periphery to extract, compare and guide the gaze. They also suggest that peripherial information is available early in a fixation to decide where to fixate next. Their data indicate that the area processed within a fixation is fairly large and contain several objects, but also that when an observer need a very high degree of detail they will fixate on the object directly.

Irwin [10] found that when observers are viewing images with modifications, they do not remember information from one fixation to another unless they carefully attend to and encode a piece of the image.

2.2 Image difference metrics

From the CIELAB in 1976 many different image difference metrics have been published. In a number of studies it has been shown that the results do not always correlate with the perceived image difference [11, 12].

2.2.1 CIELAB ΔE^*_{ab}

In 1976 CIE published the CIELAB ($L^*a^*b^*$) color space specification, with the idea of a perceptually uniform color-space. In a color space like this it is easy to calculate the distance between two colors, this is called the Euclidean distance. A sample color with CIELAB values L^*_s ,a^*_s,b^*_s and a referance color L^*_r ,a^*_r,b^*_r. The distance is given by

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \tag{2.1}$$

where $\Delta L^* = L^*_s - L^*_r$, $\Delta a^* = a^*_s - a^*_r$ and $\Delta b^* = b^*_s - b^*_r$.

Although it was designed to derive color difference for a single pair of color patches it is widely used in the graphic arts industry. Many other color difference formulas are based on the Euclidean distance.

The most common way of using ΔE^*_{ab} as an image difference metric is by calculating the color difference in each pixel and finding the mean of these values.

$$\Delta E^*_{ab} = \frac{\sum_{x=0}^{m} \sum_{y=0}^{n} \Delta E^*_{ab_{(x,y)}}}{m \cdot n} \tag{2.2}$$

where m is the width of the image and n is the height of the image. Other measures of the ΔE^*_{ab} can be the minimum value or the maximum value in the computed difference.

2.2.2 The CMC

The CMC color difference (ΔE_{CMC}) formula is based on the colorimetric principles of the CIE 1976 system. The CMC formula has acceptance in industrial color control applications [13]. ΔE_{CMC} is a modification of CIE $L^*C^*h^*$ color difference [13].

2.2.3 The CIE ΔE_{94}

The CIE ΔE_{94} [13] was developed as it became clear that the CIELAB ΔE^*_{ab} did not correlate with the perceptual color difference. In 1995 the CIE published this formula, named CIE94. This formula is based on CIE lightness ΔL^*, chroma ΔC^*, and hue ΔH^* differences.

$$\Delta E^*_{94} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*}{k_H S_H}\right)^2} \tag{2.3}$$

where k_L, k_C, k_H are scaling parameters, S_L,S_C,S_H are lightness, chroma and hue scaling functions [13]. ΔL^*, ΔC^* and ΔH^* are referred to lightness, chroma and hue differences.

2.2.4 The CIE ΔE_{00}

The CIE ΔE_{00} was published in the year 2000, because of the same problems as CIE ΔE_{94} [13].

$$\Delta E^*_{00} = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H^*}{k_H S_H}\right)^2 + R_T \phi(\Delta C^* \Delta H^*)} \tag{2.4}$$

where k_L, k_C, k_H, S_L, S_C, S_H are the same as in ΔE_{94} and R_T is an additional scaling function depended on chroma and hue [13]. ΔL^*, ΔC^* and ΔH^* are referred to lightness, chroma and hue differences.

2.2.5 Hong and Luo's hue angle metric

This algorithm for color difference is based on the known fact that systematic errors over the entire image is quite noticeable and unacceptable. The algorithm proposed by Hong and Luo [14] is based on some conjectures, these are:

- Pixels or areas of high significance can be identified and a suitable weight allocation can be found.
- Larger areas of the same color should be weighted higher.
- Larger color difference between the pixels should get higher weights.
- Hue is an important color perception for discriminating colors within the context.

The first step is to transfer each pixel in the image from L^*, a^*, b^* to L^*, C^*_{ab}, h^*_{ab}. Then a histogram based on the hue angle is computed, and sorted ascending so weights can be applied to 4 different quartiles of the histogram. The overall color difference is then calculated by multiplying the weighted hue angle for every pixel with the color difference pixel-by-pixel.

2.2.6 SSIM

The SSIM (structural similarity) index proposed by Wang et al. [15] attempt to quantify the visibility between a distorted image and a reference image (Figure 1). The algorithm define the structural information in an image as those attributes that represent the structure of the objects in the scene, independent of the average luminance and contrast. The index is based on a combination of luminance, contrast and structure comparison. The comparisons are done for local windows in the image, the overall image quality is the mean of all these local windows.

$$MSSIM(X,Y) = \frac{1}{M}\sum_{j=1}^{M} SSIM(x_j, y_j) \tag{2.5}$$

where X and Y is the reference and distorted images, x_j and y_j are image content in local window j and M indicate the total number of local windows. Figure 1 shows the SSIM flowchart, where signal x or signal y has perfect quality and the other is the distorted image.

A color extension has also been developed and tested, where each SSIM for each channel in the IPT color space were performed [12]. Then all three channels were combined with a geometrical mean.

2.2.7 S-CIELAB

The S-CIELAB model [16] was designed as a spatial pre-processor to the standard CIE color difference equations, to account for complex color stimuli such as halftone patterns [17]. As seen in Figure 2 the S-CIELAB has two goals, first it would like to apply a spatial filtering operation to the color image data, to simulate spatial blurring by the HVS (human visual system). Second, in large uniform areas, the authors would like the extension to be consistent with the basic CIELAB calculation [17]. The image data goes through a color separation where they are transformed into opponent-color space, where each of

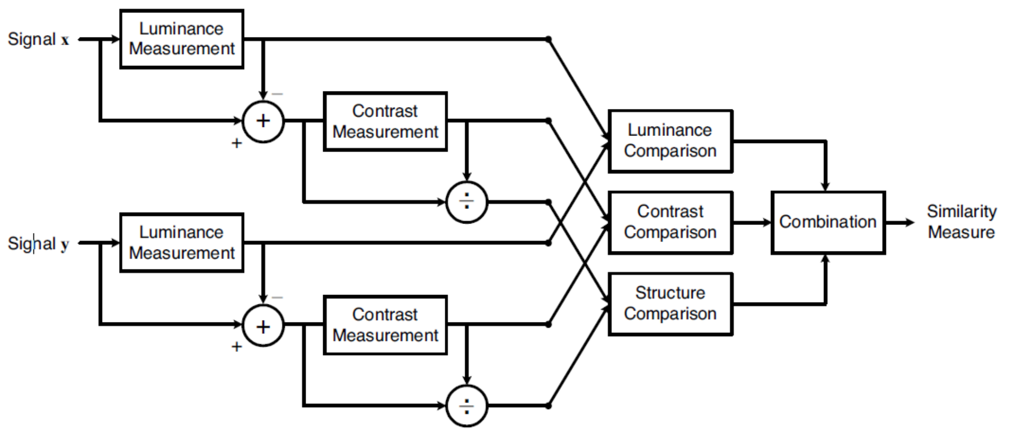


Figure 1: SSIM flowchart. Picture from [15].

these opponent-colors are convolved with a kernel. The filtered representation is transformed to CIE-XYZ representation, resulting representation includes both spatial filtering and the CIELAB processing. The difference between the S-CIELAB of the original image and its reproduction measure error of the reproduction. The difference is summarized using ΔE_s, similar to the ΔE in the conventional CIELAB.

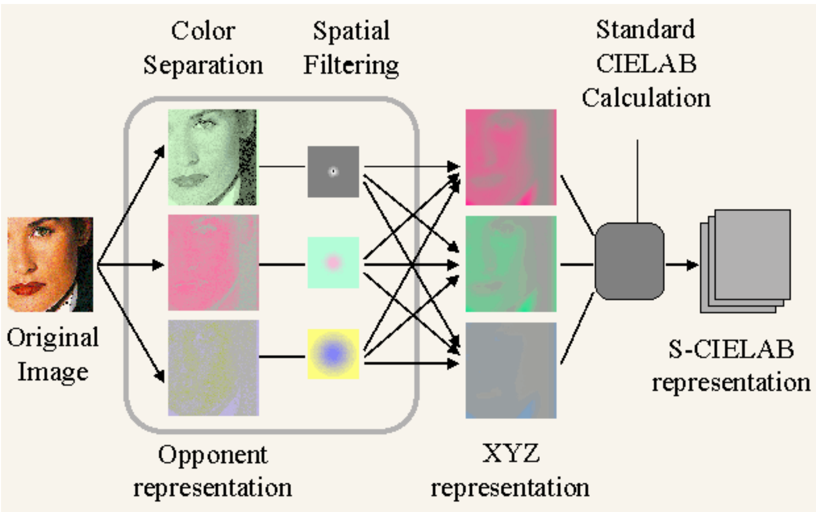


Figure 2: S-CIELAB flowchart. Picture from [18].

2.2.8 iCAM

The iCAM model was proposed by Mark D. Fairchild and Garrett M. Johnson [19]. This model was built upon previous research in many fields among uniform color space [20] because of the hue-linearity [21], the image surround importance [22], image difference and image quality measurement algorithms [23, 24]. As seen in Figure 3 this model takes the tristimulus values and transform them into RGB values using Von Kries adaptation identical to the one found in CIECAM02 [25]. Further the adopted signals are transformed into the IPT color space [20]. The adapting and the surround luminance levels are then used to allow for the prediction of various appearance phenomena. Further, the signals are transformed into JCh and with the adapting luminance information, a convertion to brightness and colorfulness predictors is performed. The quality measure is then build on the appearance correlates, and the color difference is calculated as the Euclidean distance.

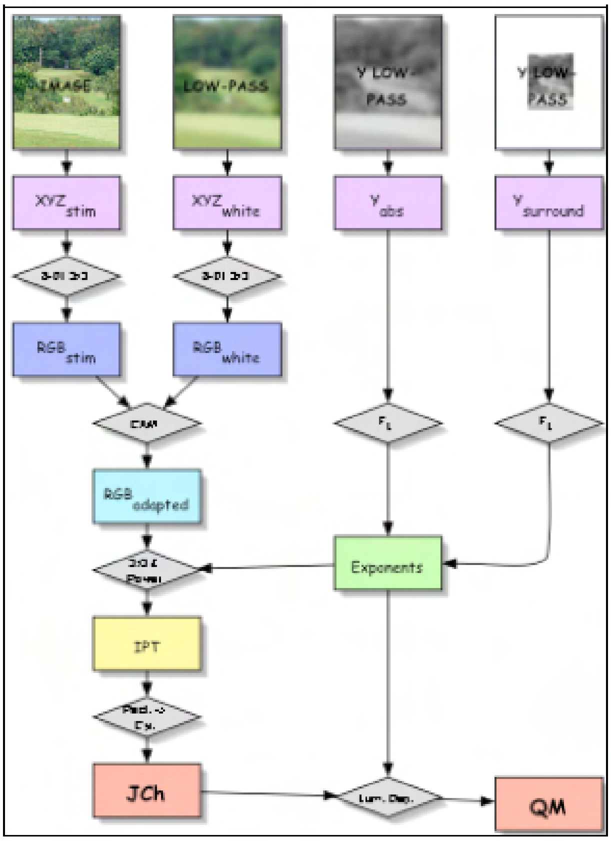


Figure 3: iCAM flowchart. Picture from [26].

2.2.9 ΔI_{cm}

This image difference metric was proposed by Morovic and Sun [27]. This metric is based on previous work by the same authors [28], where they try to understand what factors that contributes to judgments made by observers in experiments where they judge the quality of color reproduction. They proposed a seven step based image difference metric [28]. These seven steps includes the use of a 99th percentile in ΔE_{97s} from CIECAM97s2. A CSF filter(contrast sensitivity filter) is also applied, the same as in S-CIELAB. A weighting of the ΔE_{97s} with the ratio 1:2:1 for ΔJ, ΔC and ΔH. A proportion of unacceptable differences are also taken into account. The distribution of lightness differences is included. The lightness and chroma from the originals are also incorporated and at last how the spatial details have changed from the original to the reproduction.

2.2.10 Visible Differences Predictor

This is an algorithm proposed by Scott Daly [29]. The goal of the Visible Differences Predictor (VDP) is to determine the degree to which physical differences become visible differences. This is not an image quality metric, but addresses the problem of describing the differences between two images. The output from this algorithm is an image with the visible differences between the images. The VDP can be used for all image distortions including blur, noise, algorithm artifacts, banding, blocking, pixellation and tone-scale changes. There has also been developed a wavelet extension of VDP [30]. Another metric has been built on the same principles as VDP. This is called Functional Difference Predictors (FDPs) [31]. VDPs predict whether images will be visibly different, but FDPs predict whether they are functionally different, affecting the user's ability to perform a task [31].

2.2.11 Other metrics

Neumann et al. [32] proposed a perception based image metric. This metric takes into account the contrast sensitivity function and the CIELUV color space. A number of rectangles are placed in the image, in each rectangle the difference is calculated and summed up to an overall difference in the image.

2.3 Visualization of data

Babcock [33] uses a gaussian filter smoothing of fixations points to visualize the eye tracking data, both in 2d and 3d. The same 3d representation was used by Bai et al. [34] on a frequency map.

Another way to visualize the data is used by Le Meur et al. [35]. They highlighted the ROI by leaving the non fixated areas dark. The advantage by using this kind of visualization of the fixation is that it is easier to see what parts of the image that are important.

D.S. Wooding [36] states that fixation maps can be used in a predictive way to identify region-of-interest and their relative attractiveness to the eye movement system. He also states that fixation maps are an useful way of visualizing large eye movement data sets.

D.S. Wooding [37] introduced the use of fixation maps as an analysis tool for eye movement traces. The fixation maps can be used to define coverage, the amount covered by fixations, area-of-interest and similarity between observers.

2.4 Region-of-interest

Region-of-interest extraction is normally done by algorithms, but it has also been done by using an eye tracker. In research by Ukita et al. [38] gaze points have been used to extract regions in an image together with edge detection, allowing extraction of region-of-interest for both static and moving objects.

Santella and DeCarlo [39] recorded point-of-regard (POR) from an eye tracker were used to determine ROIs. These ROIs are computed using clusters of POR on a mean shift procedure. The same authors used eye tracking as an useful tool for evaluation of NPR (Nonphotorealistic rendering) systems [40].

Miyata et al. [41] performed a psychophysical experiment to verify total image quality of degraded images by a physical quality criterion (PQC) calculated from the gazing area. The conclusion here was that PQC calculated in the gazing area correlate well with the observers rating value and is not dependent on scene content.

Tillander [42] made a design and an implementation of a ROI analysis system for eye tracking. The ROIs could be automatically or manually created, and they were editable after creation. Two types of ROI-extractions were done; edge detection and color segmentation.

Ko and Nam [43] proposed a novel OOI (Object-of-Interest) segmentation algorithm for natural images that is based on the human attention and semantic region merging. Here an image segmentation together with a saliency map and saliency points are used to get an attention window. This attention window is then used with a SVM (support vector machine) to get the OOI.

Fan and Zhu [44] proposed an automatic model-based semantic object extraction algorithm by integrating object seeds with their perceptual models. This is done by a partitioning of homogeneous regions with boundaries and an edge detection.

Osberger et al. [45] used an importance map to weight errors according to the region they occurred in. The result of this was a better prediction of the locations of important areas compared to PSNR (peak signal-to-noise ratio).

Privitera and Stark [46] compared human ROI to algorithmically ROI. The overall results from this research showed a similarity between the two ways of computing ROIs.

Babcock [33] did a comparison between the regions used to make preference decisions and the fixation points in his master thesis. Babcock concluded that there is not necessarily an agreement between peak areas and the user marked region-of-interest. The main goal in this thesis was to compare different image quality evaluation methods, and the results showed a high degree of consistency. In the pair comparison task he found no tendency to fixate longer on the left or the right image, however he found that people had a tendency to fixate longer on the preferred image versus the discarded image. It was also identified that observers peak areas of attention were drawn toward faces and semantic regions. This is the same as Yarbus [47], where he showed that eyes were drawn toward ”useful or essential” areas and that we use more time on informative areas.

Le Meur et al. [35] proposed an approach to the modeling of the bottom-up visual attention. This model is based on the basics of the HVS (Human visual system).

Mannan et al. [48] argued that informative areas of the image are identified within a few seconds. In the same article they examined a set of features as contrast, luminance, high spatial frequency content and density to see if there is a correlation between these features and the fixation patterns. They did not find a correlation here, but they found that the fixation patterns between the observers were conserved. They also observed that the distribution of fixation locations is non-uniform, with a bias toward the center of the images. In work done by Buswell it was discovered that the center of the image received the most attention [33].

Babcock [5] carried out a revised experiment of Buswell's in 1935. In general no two observers had the same viewing behavior. They tended to make quick, global fixations early, and then change to longer fixations and smaller saccades as the viewing time increased.

2.4.1 Region-of-interest and image difference metrics

In research by Bando et al. [49] regions based on criteras from the observers were used to compute areas used to calculate image difference on gamut mapped images. The pixelwise difference inside the regions were computed, and pixels outside were discarded. Metrics used to compute pixel-by-pixel difference were ΔE^*_{ab}, S-CIELAB and iCAM. The conclusion by the authors was no correlation between the average pixel-by-pixel difference and perceived image quality.

Morovic and Sun [28] did research on differences in color image reproduction experiments. They discovered that more than 50% of the important errors were perceived in parts of the image rather than in the entire image. They also found out that size and location of an object within an image do not have a strong impact on whether its difference will be jugded important. In the following research they found that 80% of differences were reported due to changes in lightness (L), colorfulness (C) and hue (H) [27]. The ratio between these (L:C:H) were calculated as 1.0:1.0:0.3.

Bai et al. [34] evaluated S-CIELAB on images produced by the Retinex method by using gaze information. The average S-CIELAB color difference was weighted by the frequency map from the gazing information on the whole image and over the gazing areas. The results from this research show that frequency distribution of gazing area in the image gives important information on the evaluation of image quality.

Osberger et al. [50] presented a method for automatically determining the perceptual importance of regions in an image. The importance map created here used factors as contrast, size, shape, location and background for map calculation. This method showed great results for simple images, and the results for complex images were stated as good.

2.5 Eye tracking in research

Eye tracking equipment has been used in a number of research experiments. One of the fields are usability evaluation on web pages [51, 52], and a number of software has been developed for analyzing the results from an eye tracker [53]. The process to evaluate usability in web pages are done in several papers. [52, 54, 55] used scan paths to determine how people use webpages and another research experiment used heat maps [56]. Kukkonen et al. [57] used heat map from an eye tracking experiment to find correlation between gaze and different aspects of design evaluation. Koivunen et al. [58] found differences in gaze pattern when observers where given different tasks when perceiving product designs.

Vuori et al. [59] used eye tracking to find out if subjective image quality evaluation could be quantified. The experiment showed that image quality and instruction had significant effect on eye movement. They also concluded that eye tracking provided an interesting addition to image quality studies.

Law et al. [60] showed differences in eye movements between experts and novices in a virtual laparoscopic surgery training environment, experts were quicker and more precise than novices.

2.6 Eye tracking equipment

The eye tracking equipment is a contact free gaze measurement device, the eye tracker used can be seen on Figure 4. The software allows online gaze position computation, real-time visualization, online fixation analysis, and digital output for control purposes [61]. The results from the eye tracker can be analyzed with a 3rd party software.

- Sampling Rate : 50/60 Hz
- Tracking Resolution, Pupil/CR : 0.1 deg. (typ.)
- Gaze Position Accuracy : 0.5 - 1 deg. (typ.)

- Operating Distance Subject-Camera : 0.4 - 1.0 m
- Head Tracking Area : 40 x 40 cm at 80 cm distance

Figure 4: SMI remote eye tracker with infrared lightsource on top.

2.6.1 Background understanding

Humans make over 100 000 eye movements a day [2, 33]. In the context of viewing static images, eye movements can be described as a sequence of fixations and saccades. Fixation is when the eye has paused on a particular position in the image, saccade is the period when the eye moves from one fixation to another [33]. In a free viewing human make several eye movements per second, in general these movements are 3-4 saccadic eye movements [2, 46].

SMI Remote Eyetracking Device (RED) is a dark pupil system. The eye is illuminated with an infrared (IR) lightsource from an angle with an IR camera. The pupil will absorb most IR light and appear as a high contrast dark ellipse [62]. From this ellipse the center can be found using a set of algorithms. Along with this the cornea reflex (CR) is tracked, because of the IR light directed to the eye a reflection will arise and this can be tracked. The advantage by doing this is that the head position is relative to the camera, based on this along with the pupil location the gaze point in the stimulus can be determined [62].

3 Experimental setup

3.1 Experimental enviroment

This section explains the setup for the experiment, the eye tracker, instructions and everything connected to the setup.

3.1.1 Experiment setup

The psychophysical experiment was done on a calibrated CRT (cathode ray tube) monitor, LaCIE electron 22 blue II, hooked up to a computer with an Intel Pentium III (731MHz) processor and 384MB ram. The experiment room was painted gray with no windows and the tables were also gray. The setup was similar to the setup used by Cui [63] regarding the room.

The resolution of the CRT monitor was set to 1280x1024, and the eye tracker software was configured to match this resolution.

All observers were given the same instructions (see Section 3.3.1), this was to assure similar results across different observers. Several studies have shown that different instructions give different eye movements [59, 64]. It was also stated to the observers that they should ask if something were unclear or difficult to understand, then further instructions would be given.

3.1.2 Viewing distance

If the item to be viewed is located to close to the participants, both the ocular and the lenticular muscles begin to work, and therefore requiring more effort to see the item clear [65].

The recommended viewing distance by Bhattacharya et al. [65] is between 50 and 75 cm, it is also recommended to have the monitor placed to give a downward gaze angle similar to the study by Von Noorden [66]. In [67] the average resting position of the eyes was found to be 59 cm. It is also stated that the resting position becomes further away as the age increases.

For this experiment the observer was seated approximately 80 cm from the screen [68], and approximately the same distance from the eye tracker giving a head tracking area of about 40x40 cm [61]. 80 cm from the screen also gives a viewing angle of approximately 27$\times$21 degrees. The participants were seated so that they had a downward gaze angle. Several studies done by Von Noorden showed that participants had head movements even at small movements of the eyes [68]. According to Von Noorden a downward gaze angle results in smaller head movements [66], this will decrease errors in the data set. The chair used had 4 legs, armrests and backrest, this type of chair was choosen to minimize observer movement.

Time consumption for the experiment was calucated to a mean of approximately 30 minutes. With an experiment time this long a normal and not to stressing situation for the eyes are important.

Image name	Image size
TORE	340x515
GIRL	340x466
JP	340x453
CARTOON	260x348

Table 1: Image resolutions. Images were resized to match screen resolution and the eye tracker software.

3.1.3 Light conditions

According to the iViewX manual light changes should be avoided [62], therefore the light conditions were not altered during the experiment. Ambient light is recommended by SMI, due to better accuracy. Complete darkness will effect the accuracy of the eye tracking.

The light condition during the whole experiment was dimmed. The luminance in the room was measured using a Sekonic Flashmate L-308s light meter [69]. This gives the exposure value (EV) as a measure of luminance. The CIE guidelines for gamut mapping evaluation states that the light conditions should be lower than 64 lux [70]. The relationship between lux and EV is [71, 72]

$$E = 2.5 \cdot 2^{EV} \tag{3.1}$$

The measured exposure value in the room with dimmed lighting and the screen turned on was 2.1 on the back wall and 2.8 in front of the screen, from this the E is calculated to approximately 11 lux on the back wall and approximately 17 lux in front of the screen.

3.1.4 Images

4 different images (Figure 5) have been used in this experiment. Girl reproduced with permission of Grafisk Assistans [73], Tore is reproduced with permission from "Se og Hør", JP is from ISO [74] , Cartoon is made by Trond Viggo Bjerke and reproduced with his permission. The three first images have been used in similar experiments before, and the last was choosen due to the simplicity of the image.

(a) Girl (b) Tore (c) JP (d) Cartoon

Figure 5: Images used in the experiment.

All images have been reproduced with different changes in lightness. Each scene has been altered in 4 ways globally and 2 different regions, each of the two regions are then altered 2 ways locally, resulting in 8 different reproductions for each scene. The images were then altered in lightness with MATLAB 7 and the monitor profile was applied.

Changes were made in the CIELAB color space only in the L^* channel in 16 bit, the global changes were plus and minus 3 and 5 ΔE^*_{ab}. After this the images were changed to 8-bit in order to apply the monitor profile. All images were saved as 16-bit Portable Network Graphics (PNG). This procedure introduces some quantization noise to the images, but not to a significant degree.

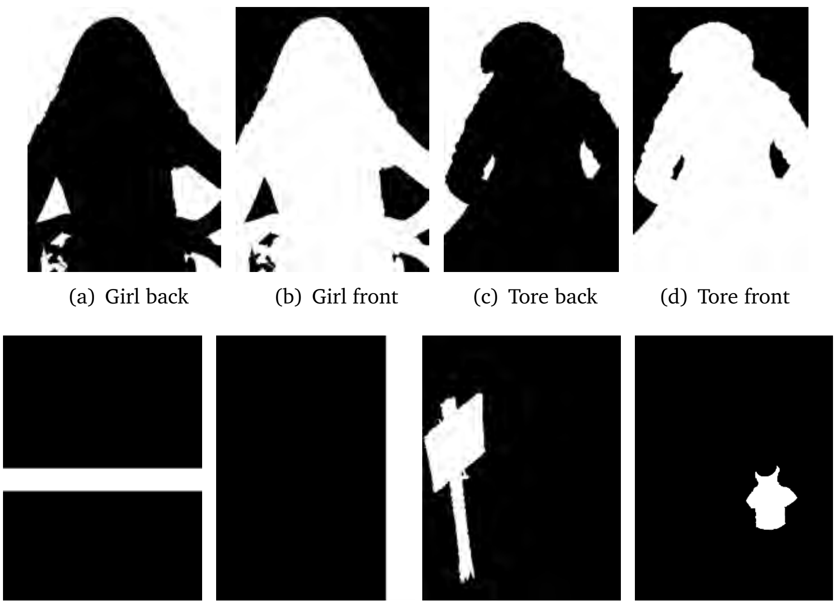
Reference image

The reference image is the original tiff resized to match the size chosen (Table 1). The monitor profile was then applied and the image saved as a 16-bit Portable Network Graphics (PNG).

Regions

In the images with only changes to regions, the regions have been altered in lightness to the same degree as the globally changed images. Resulting in that the regions have been altered to the same degree as the same region for the globally changed images.

The different regions altered (Figure 6) were front and back for the Girl (Figures 6(a) and 6(b)) and Tore (Figures 6(c) and 6(d)) scene. This was to see if changes made to back or front are rated different. On the other hand in JP scene 2 different stripes were selected (Figures 6(e) and 6(f)), one horizontal and one vertical to see if these are visually different. In Cartoon scene two small regions (sign in Figure 6(g) and tshirt in Figure 6(h)) were changed. This was done to check if modification to semantic regions are perceived as important. The white areas in Figure 6 were altered and the black were not. An example of back altered with 3 ΔE_{ab} can be found in Figure 7. The area changed are from 3% to 68% of the whole image (Table 2).



(a) Girl back (b) Girl front (c) Tore back (d) Tore front

(e) JP stripe (f) JP stripe2 (g) Cartoon sign (h) Cartoon t-shirt

Figure 6: Regions, only white pixels have been altered.

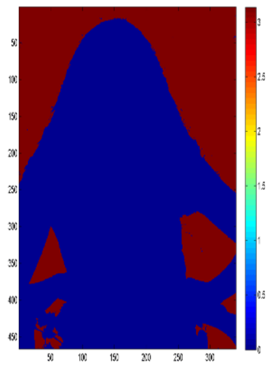


Figure 7: Back region in Girl altered with 3 ΔE^*_{ab}, blue is unchanged while the dark red indicate 3 ΔE^*_{ab} difference.

Image name	Percentage
GIRL BACK	32 %
GIRL FRONT	68 %
TORE BACK	28 %
TORE FRONT	72 %
JP STRIPE	8 %
JP STRIPE2	14 %
CARTOON SIGN	10 %
CARTOON TSHIRT	3 %

Table 2: Percentage image area altered in regions.

3.1.5 QuickEval 2.0

In this project paired comparison has been used as a psychophysical method, this is also recommended by Cui [63] for these types of experiments. The program used was QuickEval 2.0, this has earlier been used in another paired comparison experiment [49]. The original Java software [75] only allowed for showing 2 images at once, left and right side. Therefore the program was modified by the author to fit 3 images, one reproduction on each side and the original image in the middle (Figure 8). A freeview task was

Figure 8: QuickEval 2.0 screenshot, original in the middle and 3 reproductions on each side. Observers were told to choose the image most similar to the original.

also included in the first part of the program and a gaze marking task in the last part, where 4 images were showed at once (Figure 9). In the freeview task observers where told to look freely at images, while in the gaze marking they where asked to look at the areas important for their choice. Code for communicating with the eye tracker software

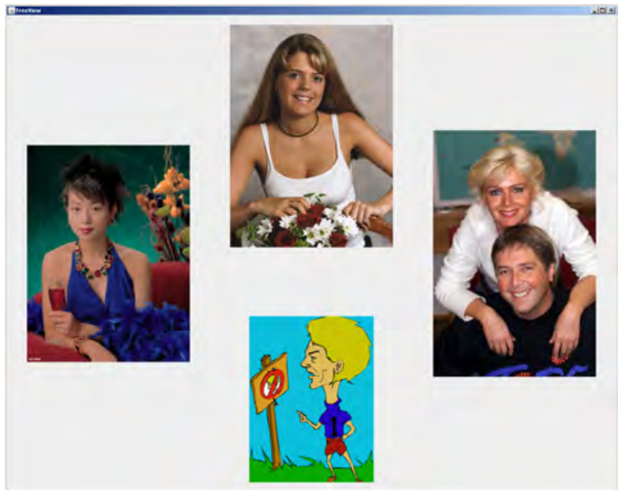


Figure 9: QuickEval 2.0 freeview and gaze marking screenshot.

iViewX was developed by the author (Code in appendix B.1). This automates the process with increasing set number, used to organize the recorded data, for each image and also sending the names of the images shown including which image the observer preferred. The QuickEval program calculates the z-scores for the scenes and the confidence intervals. The 95% confidence intervals are calculated in the same way as proposed by Morovic [76].

$$\overline{X} \pm 1.96 \frac{\sigma}{\sqrt{N}} \tag{3.2}$$

where N is the size of the sample from which $\overline{X}$ was calculated. For this experiment $N =$ number of observers $\cdot 2$, because each image is shown twice for consistency.

When an observer gets one pair of images, the center image is the reference and the two sides have one reproduction each. The observer left-clicked on the image he or she preferred and a new pair was automatically shown right away. This will continue until all pairs are shown, then the observer is shown a button before moving on to the next scene. The observer will always get 2 different reproductions on each side.

3.1.6 Experiment workflow

The experiment is divided into several tasks (Figure 10). The dominant eye is first found by using the "Porta test" (see Section 3.2.1) and in the second part the observers were given instructions (Section 3.3.1). The third part is the freeview, where the observers spend as long as they like looking freely at the 4 images.

The next part is the pair comparison with the 4 different scenes, where observers rate the images. The fifth part is the gaze marking where observers look at the areas important for their decision in the pair comparison. The last part is the questionnaire including marking of important regions on paper. A complete workflow of the experiment is found on Figure 10.

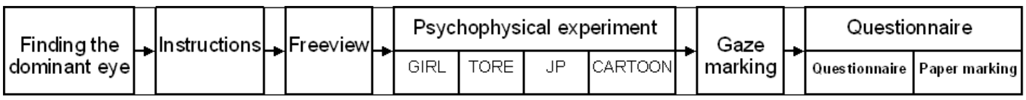


Figure 10: Experiment workflow.

3.2 Eye Tracker

3.2.1 Finding the dominant eye

Approximately 97% of the population has a visual sighting dominance, which enables them to use the same eye for primary vision. 65% of the population are right eye dominant, and 32% are left eye dominant [77, 78]. To determine whether an observer is right or left eye dominant a simple test called the "Porta test" has been used [77, 79].

- The observer is asked to point at a distant object with an outstretched arm, using both eyes.
- While still pointing, the observer is asked to close one eye at the time.
- The eye that sees the finger pointing directly at the object is the dominant.

3.2.2 Ensuring good eye tracking results

In order to get a good result from the eye tracker some measures were taken [78, 80].

- The observers were asked to minimize head and body motion.

- The eye tracker was set up relative to the observer's position to avoid the lower eyelid to obstruct the view of the iris and pupil.
- Observers were over-recruited in case of calibration failure, loss of eye, technical problems and similar problems.
- Observers with glasses and contact lenses were limited during recruitment.
- Observers dominant eye was preferred.

3.2.3 Eye tracker calibration

The calibration of the eye tracker is very important to relate the observer's point-of-regard to the location at the screen. Poor calibration can result in a mismatch between the point-of-regard and the corresponding location at the screen. In this project a 9 point calibration routine with corner correction has been used, this has been used in other eye tracking experiments [33, 59]. The 9 point calibration is also recommended by SMI [62]. The eye tracker was calibrated according to each observer before commencing the experiment. This must be done because each observer has a unique shape of the eye [81].

3.2.4 Eye tracker data

The data gathered by the eye tracker was exported to a text file for further analyzing in MATLAB. The exporting was done with IDFconvert, a software received with the eye tracker. This software converts the idf file from the eye tracker software, iViewX, to a plain text file. The converted file (Table 3) contains a header with information about the calibration, head distance, date etc. The rest of the file contain the data gathered by the eye tracker.

Time	Type	Set	R Raw X	R Raw Y	R Dia X	R Dia Y	R CR1 X	R CR1 Y	R POR X	R POR Y	Timing	Latency
3091825464	SMP	1	182.55	142.40	21.52	1455.00	193.47	153.88	878.60	527.77	0	22594
3091845489	SMP	1	183.18	143.37	21.75	1486.00	194.10	154.93	881.25	526.38	0	22700
3091865463	SMP	1	183.78	143.08	21.51	1454.00	194.53	154.60	875.93	525.33	0	22522
3091885491	SMP	1	190.92	142.78	20.13	1273.00	199.10	154.84	713.45	494.99	0	22653
3091905464	SMP	1	196.09	141.28	21.85	1500.00	201.81	153.90	574.43	485.43	0	22692

Table 3: Eye tracker sample data.

The first column "time" contains the time for when a fixation was made in microseconds. The time used for one fixation can be calculated by taking the timestamp for the next fixation and subtract the timestamp from the current fixation (Equation 3.3) for valid data points. When the observer is blinking or looking outside the screen the values are not valid and cannot be used. Therefore if there is a blink the coordinate values are set to 0 by iViewX and if the observer looks outside the screen the values are negative, below coordinates (0,0) or above coordinates (1280,1024).

$$T_{fixationtime} = T_{nextfixation} - T_{currentfixation} \tag{3.3}$$

The next column indicates if the row contains a sample (SMP) or a message (MSG). If it contains a sample it will be like Table 3. If the second column contains MSG, the following on the same row will be a message. This can be image names or messages about recording etc. The third column indicates the "set number", this is used to differentiate between scenes, tasks etc. In this case the set number was incremented each time the observer changed image. This allows for extracting the time and pixels fixated for each image or scene shown. R RAW X and R RAW Y indicates the raw data for the x and y component, R Dia X and R Dia Y are the diameter position for the x and y component. R CR1 X and R CR1 Y are the first corneal reflex of the eye, where x and y are the coordinates. R POR X and R POR Y is the point of regard (x and y coordinates), the pixel value on screen. The two last colums indicates the timing violation and latency.

3.2.5 Visualization of data

To visualize the data several methods can be used. Babcock [33] used a 3d visualization with a gaussian filter, similar to the one on Figure 11. This representation of the data makes it easy to see high differences in the map, but more diffucult to see exactly where the different heights belong on the image.

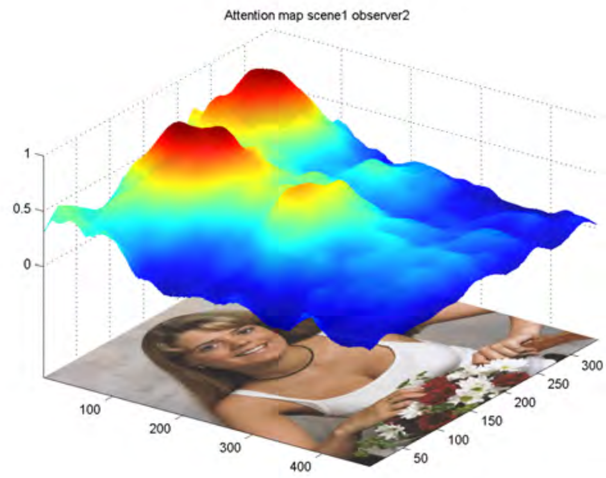


Figure 11: 3d visualization of eye tracker data.

A 2d visualization is also possible (Figure 12) with a colormap indicating differences, this makes it more difficult to see changes in height, but a little easier to see where the observers have been looking.

Another way of showing the same data with a gaussian filter is by a 2d representation with a colormap and a background image (Figure 13). This way of showing the results make it a little bit more difficult to see height differences but it is easier to see where the points belong because of the background image.

Focus map is also one way of showing the data. This method applies a black layer over the image and revealing more of the content the more fixations the observers have to one point (Figure 14). By using this visualization it is easy to see what regions the observers have been fixating on, but we are missing details in difference levels.

We have decided to use the 2d representation with and without the background image, this is mainly because it is easier to compare maps and they are more precise when determining the observers regions-of-interest.

3.3 Instructions

3.3.1 Instructions before the experiment

Before the experiment the observers received instructions, these were more complementary than those during the experiment.

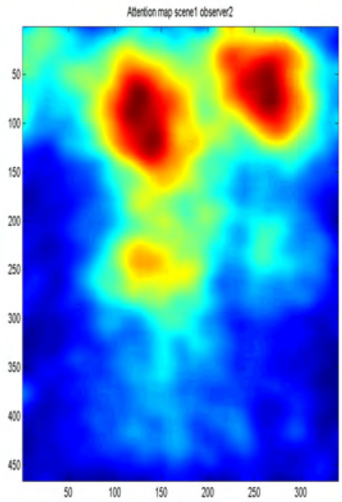


Figure 12: 2d representation of eye tracker data.

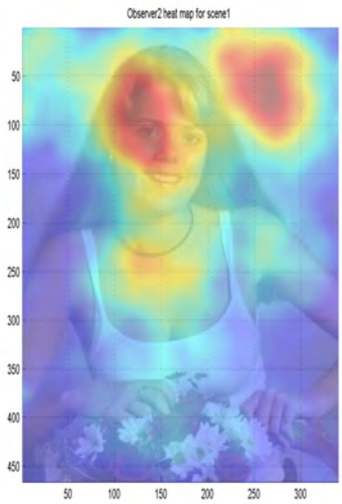


Figure 13: 2d representation of eye tracker data including background image.

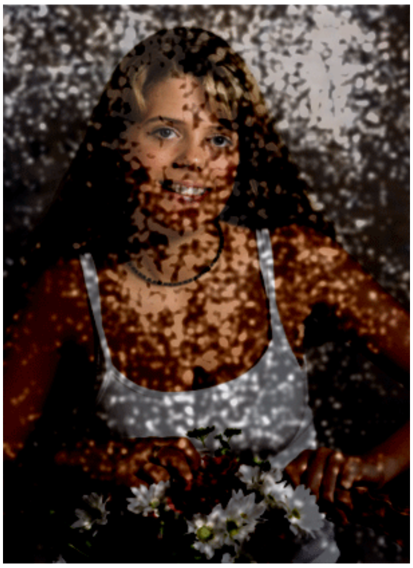


Figure 14: Focus map example.

> ”For the first part of the experiment, you will be shown 4 images. You are free to look at these for as long as you want, when you are done close the window with the ’X’ at the top corner.
>
> After doing this the second part will begin. Here you will be shown 3 images at once, the middle image is the original and the right and left image are reproductions. Your task is to choose the image most similar the original. To choose left-click once with the mouse button, when this is done a new pair will be shown.
>
> The last part is similar to the first part, further instructions will be given before that part begins. When you are done with this part close the window with the ’X’ at the top corner.
>
> In order to get the most correct results from the eye tracker please minimize your body and head movements and keep your eyes on the screen. ”

After the observer had read the intructions, he or she was asked if they understood the instructions. If these where not understood in-depth explanation was given the observer.

3.3.2 Instructions during the experiment

In the program the observers got instructions during the experiment. For the first part (freeview):

> ”You will be shown 4 images, when these appear you are free to look at these for as long as you please. When done with the free viewing, close the window and proceed to the next section.”

For the second part (the pair comparison):

> ”3 images will be shown, the middle image is an original image and the side images are reproductions. Left click once on the image most similar the original and a new pair will be shown.”

For the last part (gaze marking):

> ”You will now be shown 4 images, look at the regions that are important to you. Close the windows when you are done.”

3.3.3 Questionnaire

After the experiment the observers were asked to fill out a questionnaire (Appendix A). The observers were also asked to mark regions important to them during the evaluation of the scenes. The 4 images used in the experiment were printed in black and white and the observers marked important regions with a pen.

4 Experimental results

This chapter gives an overview of the experimental results from the psychophysical experiment and questionnaire results.

4.1 Questionnaire results

A total of 25 observers were recruited from the school with an age range from 20 to 38 years, with a mean of 24. The observers had different background, some had studied color science and were well familiar with psychophysical experiments, while others had not participated in experiments similar to this before.

All observers were asked to fill out a questionnaire (Section 3.3.3) after the experiment to map differences in background knowledge. From this we can see that 56% of the observers are experts, these observers have studied or have knowledge in the field of color (Figure 15). 24 % of the observers had participated in psychophysical experiments before, while 76% participated for the first time (Figure 16(a)). Only 8% had participated in eye tracking experiments before, and for the rest this was a first encounter with eye tracking (Figure 16(b)). 80% of the observers recognized the man on the second image (Figure 5(b)), and 20% did not recognize him (Figure 17(a)). The man on the image is a norwegian tv celebrity, who has his own popular show on norwegian tv.

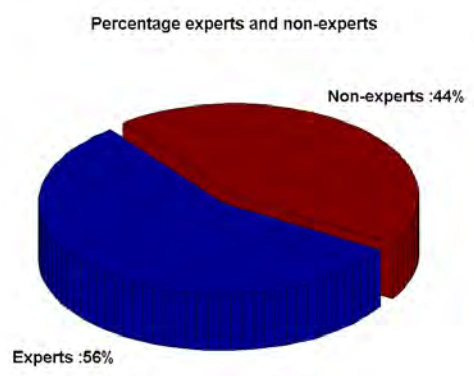
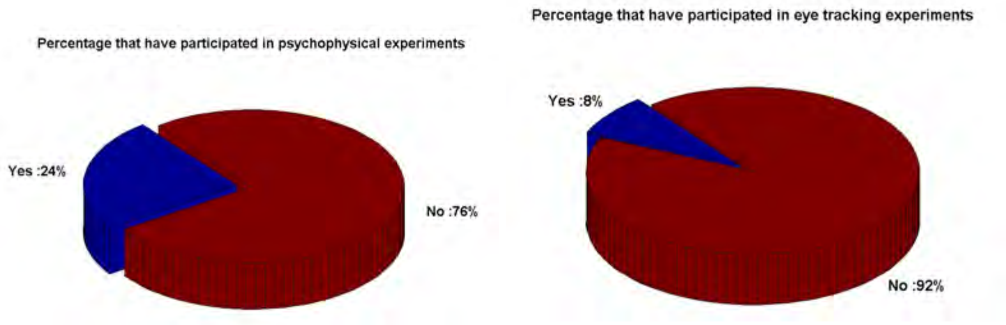


Figure 15: 56% experts and 44% non-experts participated in the experiment.

All observers were also asked if they had seen the test images before (Figure 17(b)). For the Girl (Figure 5(a)) and Tore (Figure 5(b)) image approximately 24% had seen these images before, this is probably because these have been used in other experiments earlier. 16% had seen the JP (Figure 5(c)) image before, this has also been used in experiments before. For the last image, Cartoon (Figure 5(d)), only 1 person(4%) had seen this before.

4.2 Psychophysical experiment results

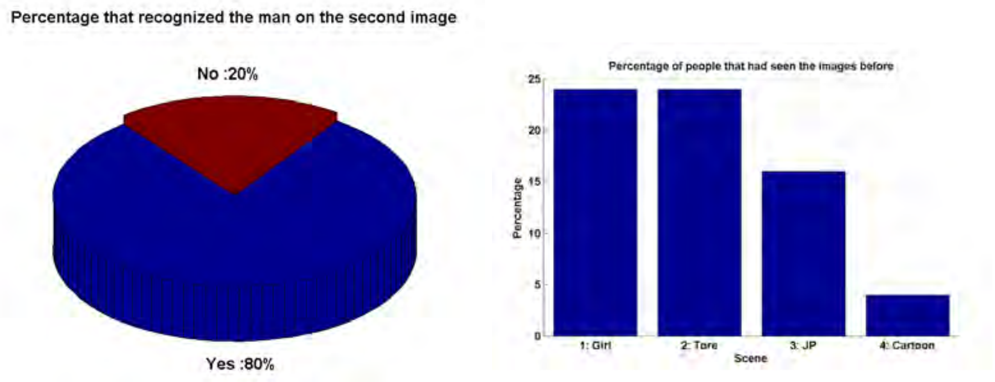
The results from the psychophysical experiment can be seen in Figure 18 and in Table 4. Figure 18 shows the mean results over the 4 scenes for the images with lightness



(a) Psychophysical experiment.

(b) Eye tracker experiment.

Figure 16: Participated in similar experiments.



(a) Recognized the man.

(b) Seen images before.

Figure 17: Recognized or seen images.

minus 3 ΔE_{ab} (hereby LM3), minus 5 ΔE_{ab} (hereby LM5), plus 3 ΔE_{ab} (hereby LP3) and plus 5 ΔE_{ab} (hereby LP5). From the results we can see that LM3 and LP3 are not distinguishable, the same with LM5 and LP5. But the observers can see a big difference between lightness 3 ΔE_{ab} and 5 ΔE_{ab}, even though there are only 2 ΔE_{ab} difference. The other 4 changes made to the images are different for the different scenes, therefore these are not discussed here.

Image	LM3	LM5	LP3	LP5
Z-score	0.1265	-0.6412	-0.0019	-0.5683
Rank	1	4	2	3

Table 4: Overall z-score global changes. Higher number the more preferred this image is by the observer.

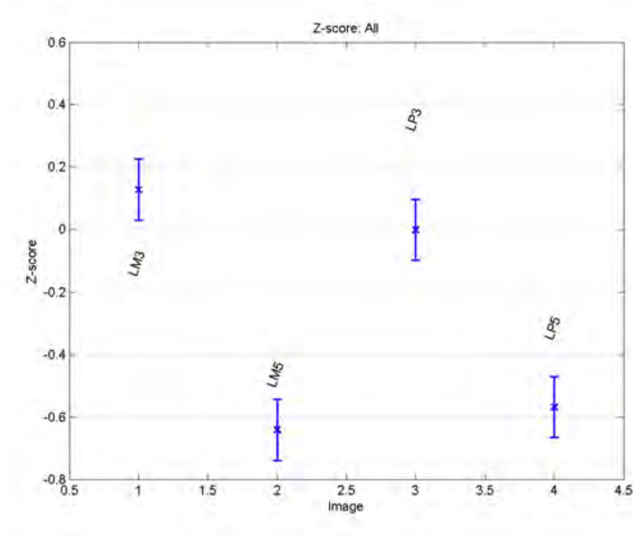


Figure 18: Z-score for the first 4 alterations. LM3 and LP3 are rated similar, while LM5 is rated similar to LP5.

4.2.1 Girl

For the girl scene, we can see that BACK LM3 (background lightness minus 3 ΔE_{ab}), BACK LP3 (background lightness plus 3 ΔE_{ab}) and FRONT LM3 (front lightness minus 3 ΔE_{ab}) score the highest, and are not distinguishable for the observers. LM3, LP3 and FRONT LP3 scores approximately the same, while LM5 and LP5 scores significantly lower than the rest. From the results shown in Figure 19 and Table 5 small lightness changes in positive direction in the background scores higher than global changes and lightness plus in the front region. A background with more lightness is prefered by the observers over a foreground with more lightness.

Image	LM3	LM5	LP3	LP5	BACK LM3	BACK LP3	FRONT LM3	FRONT LP3
Z-score	0.0490	-0.8154	0.0838	-0.6432	0.4552	0.5306	0.3018	0.03835
Rank	5	8	4	7	2	1	3	6

Table 5: Z-score Girl scene.

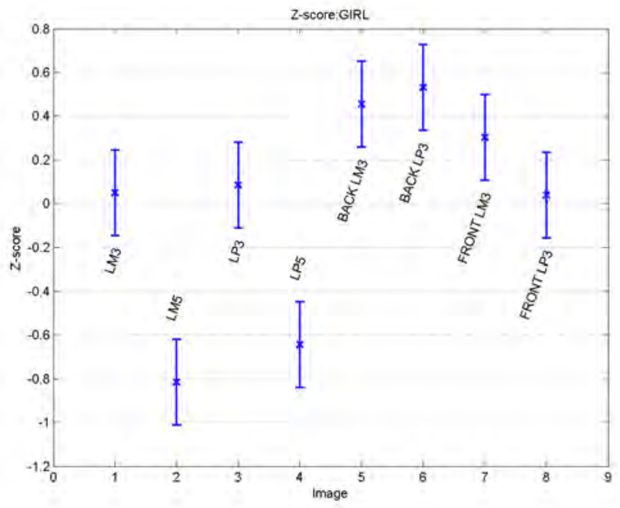


Figure 19: Z-score for Girl scene.

4.2.2 Tore

For the Tore scene BACK LM3 scores the highest (Figure 20 and Table 6), but cannot be distinguishable from the BACK LP3, FRONT LM3, FRONT LP3 and LM3, but both BACK images (BACK LP3 and BACK LM3) scores higher than LP3. Observers preferred the images with brighter or darker background over the brighter global image. Both the LM5 and LP5 score significantly lower than the rest.

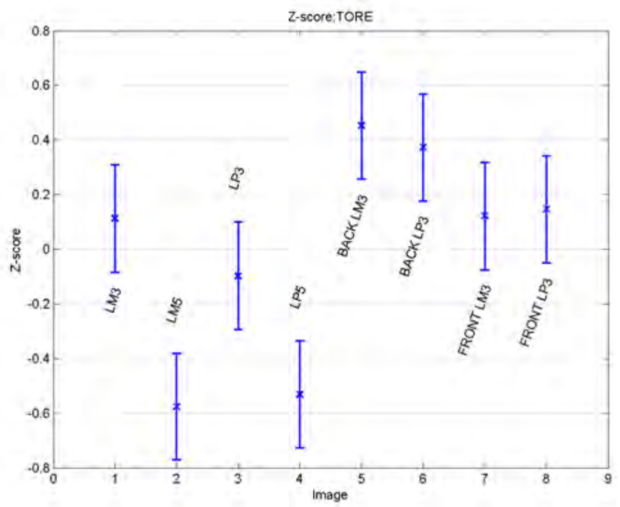


Figure 20: Z-score for Tore scene.

4.2.3 JP

For the JP scene stripe2 LP3 scores highest but cannot be distinguished from the stripe2 LM3, LM3 and LP3 (Figure 21 and Table 7). Stripe LP3 has the lowest score, but this can not be distinguishable from the stripe LM3, LM5 or LP5. In this scene observers preferred

Image	LM3	LM5	LP3	LP5	BACK LM3	BACK LP3	FRONT LM3	FRONT LP3
Z-score	0.1123	-0.5756	-0.0972	-0.5315	0.4524	0.3723	0.1219	0.1455
Rank	5	8	6	7	1	2	4	3

Table 6: Z-score Tore scene.

the low global changes and the vertical stripe on the right side of the image, while the higher global changes score lower together with the horizontal stripe in the middle of the image. The horizontal stripe scores low even though the changes made to the image are small, this is probably due to the high visibility and the placement in the center of the image.

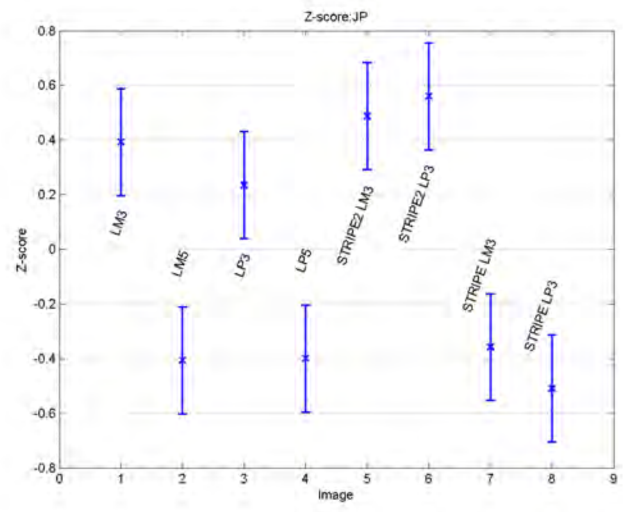


Figure 21: Z-score for JP scene.

Image	LM3	LM5	LP3	LP5	STRIPE2 LM3	STRIPE2 LP3	STRIPE LM3	STIRPE LP3
Z-score	0.3921	-0.4059	0.2339	-0.3995	0.4867	0.5597	-0.3575	-0.5095
Rank	3	7	4	6	2	1	5	8

Table 7: Z-score JP scene.

4.2.4 Cartoon

For the Cartoon scene Tshirt LM3 got the highest score, but this is not distinguishable from Tshirt LP3, SIGN LM3 or SIGN LP3. LM3 and LP3 scores lower than the previous, and LM5 and LP5 scores significantly lower than these again (Figure 22 and Table 8). With these results observers tend to prefer images with a low amount of changes only applied to regions of the images, while the larger the difference is globally the less the observers prefer them. The observers follow the rank of the percentage of changed regions (Table 2).

4.2.5 Overall observations

The overall results show that region changes are rated generally better than global changes. In both Girl and Tore scene the reproductions with modifications only in the background got the highest mean score, these are also the smallest regions. In the JP scene, the region with the smallest percentage of modification was rated the worst, but

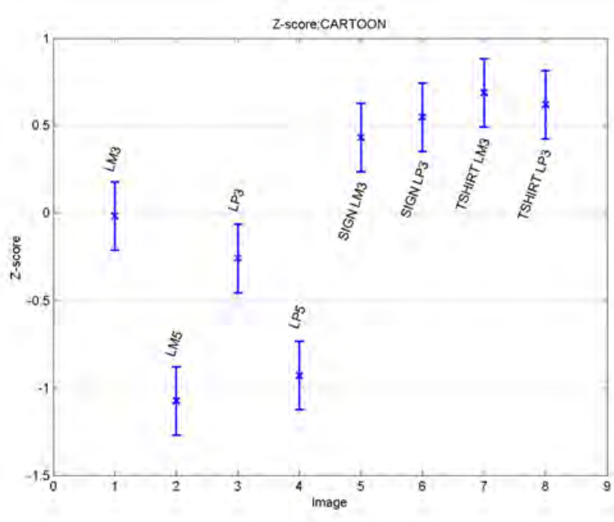


Figure 22: Z-score for Cartoon scene. Local changes are rated better than global changes.

Image	LM3	LM5	LP3	LP5	SIGN LM3	SIGN LP3	TSHIRT LM3	TSHIRT LP3
Z-score	-0.0188	-1.0739	-0.2611	-0.9292	0.4307	0.5469	0.6869	0.6183
Rank	5	8	6	7	4	3	1	2

Table 8: Z-score Cartoon scene.

the least visible change (Stripe2) was rated the highest. In the last scene the smallest change was rated the best, as in the two first scenes. It is clear that observers tend to prefer small changes, as long as they are not too visible.

5 How do we look at images

This chapter deals with how human look at images. Information about how we look at images can be valuable in image processing, gamut mapping, general modifications to an image and so on. We look at differences between observer stated regions, a freeview task, gaze marking important regions and a pair comparison task. Within each of these tasks differences between user groups, as expert and non-experts, are also analyzed.

5.1 Image analysis

An analysis of how the observers looked at the images in the different tasks are given in this section.

5.1.1 Eye tracker: freeview

Here the eye tracker map from the freeview has been computed, these can be seen in Figure 23. This map was based on the frequency map from the eye tracker [34], which was calculated as

$$F(x,y) = \left(\frac{T(x,y)}{N(x,y)}\right)_{\mathrm{NoI}} \tag{5.1}$$

where

$$F(x,y) \in [0,1]$$

The total time an observer used on each pixel was divided by the number of times the observer fixated on that pixel. The result was then normalized by the maximum value in the map [37], NoI in the formula means normalization. A gaussian filter with a width of 35 pixels and a height of 35 pixels was then applied to the map to even out differences [33, 34, 37] and to simulate that we look at an area rather than one particular point [1, 82]. An overall map was found by summing the map for each observers and dividing by the number of observers.

Girl

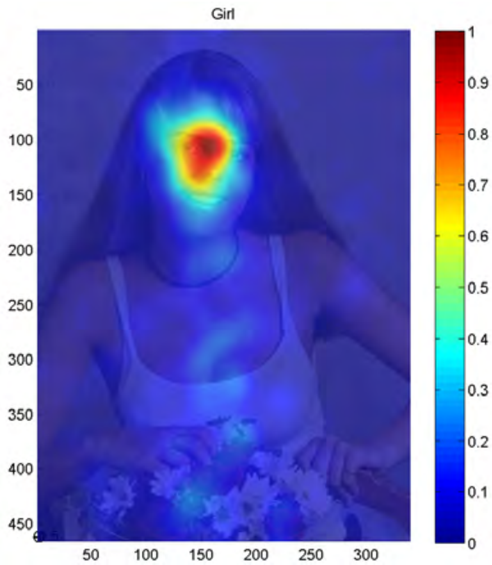
In the girl scene the face receives the most attention, the center of attention is located between the eyes. We see some fixations on the flowers in the bottom of the image and on her torso.

Tore

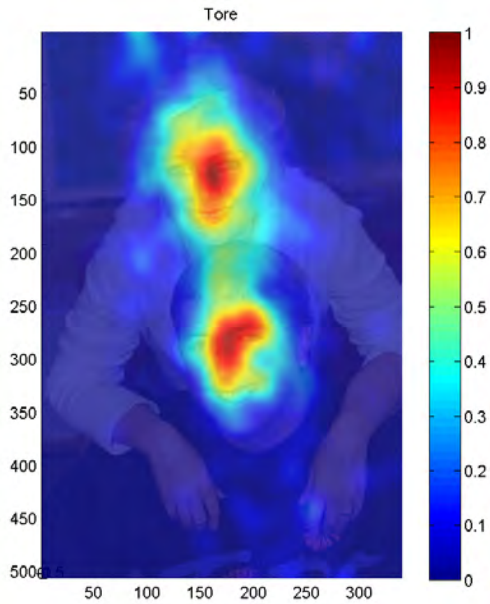
In this scene the two faces are the areas where the observers fixate, both faces are equally fixated. Some minor fixations to background is present, this could be because of the map in the background. The hands of the girl also receive some attention.

JP

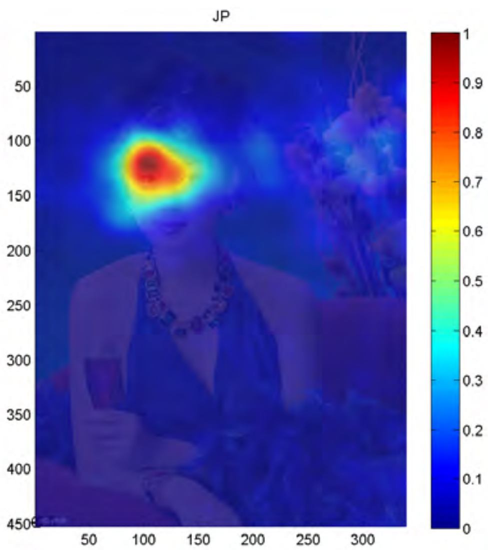
The face receives the most attention in the JP scene, center of attention is located near the girl's right eye. The flowers on the right of the image receives some attention by the observers. The area below her shoulders are more or less ignored by the observers.



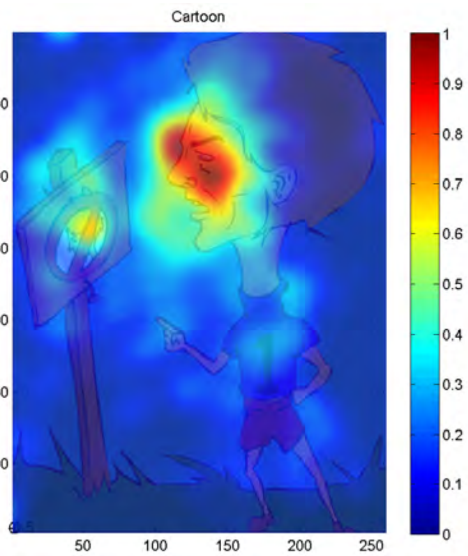
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 23: Eye tracker map from freeview task.

Cartoon

The face of the character is the center of attention in the Cartoon scene. The sign is the area receiving the most attention beside the face. Some fixations are also found on his finger and tshirt.

Overall observations

The face is clearly the most important in a freeview task, while other informative areas receive only a portion of the fixations.

5.1.2 Eye tracker: pair comparison

The frequency map from the pair comparison experiment (Figure 25) is analyzed in this section. This map was made as mentioned in Section 5.1.1. The three different images (left, center and right) were extracted and added to one overall map as seen on Figure 24.

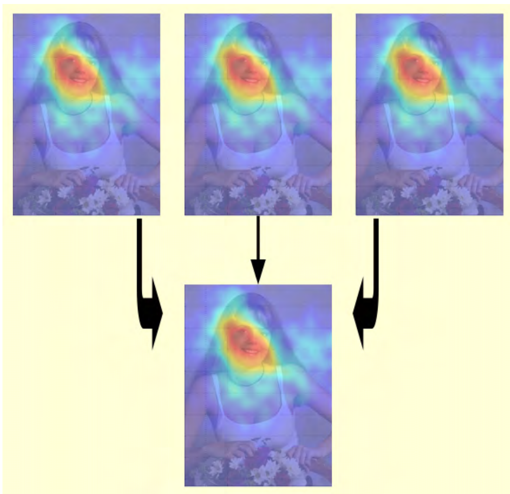


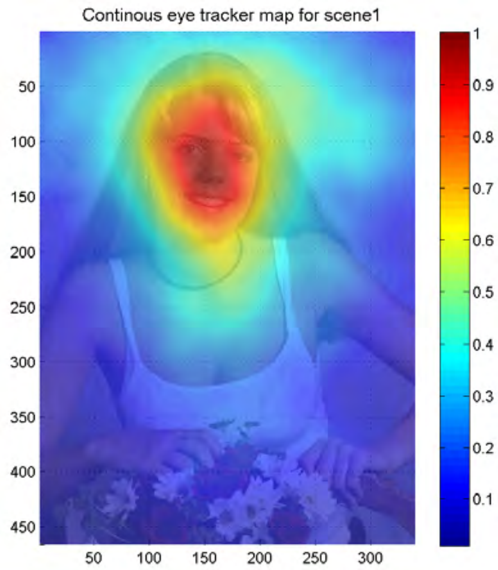
Figure 24: Workflow from 3 images to 1 freqeuncy map. Data points from the eye tracker is taken from all 3 images and merged into 1 image.

Girl

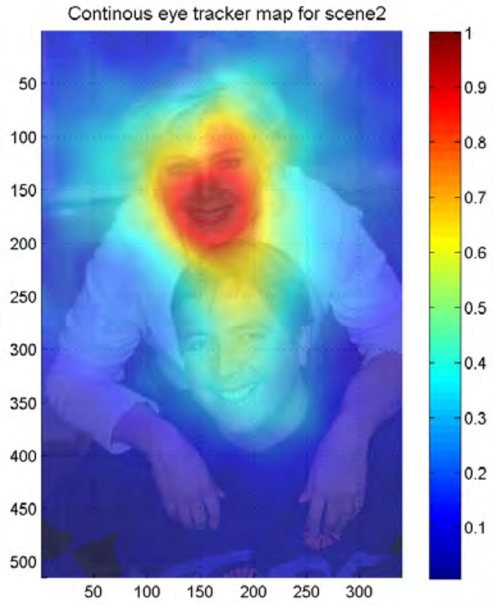
The face is the main area of attention in the Girl scene. The forehead and nose area down to the mouth receives the most fixations. The other areas of the face also get a high concentration of fixations. The face is the most informative region on the image, and is the center of attention. We do also see that the forelock on her forehead also get a high number of fixations, this area draws the most attention in the hair. Observers have also fixated on the background just right of the head, the larger and uniform part of the background. The chest and her left shoulder get a number of fixations, indicating that some observers used this area when they looked for differences betwen the reproductions. The lower end of the background, the flower bouquet and hands get little attention by the observers.

Tore

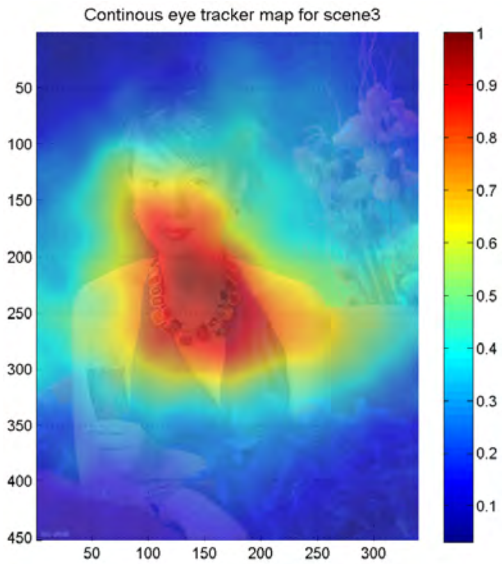
The face of the girl draws most attention in the Tore scene, the area around the mouth and nose are the most fixated. Some areas of her hair is also important. The guy's face is not as important as the girl's face. Most of the fixations here are located in the center of



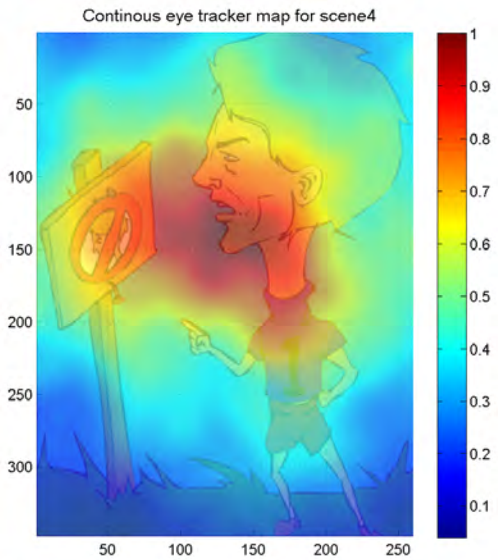
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 25: Eye tracker regions from the pair comparison experiment.

the face and up to his forehead. Semi important areas is the background on the top left and some areas of the girl's sweater. We can see that the background on the right and area beneath the guys face only have a small number of fixations.

JP

In the JP scene the neck gets the most fixations, these regions are located between the two main objects. The face as a natural attention point, and the very visible horizontal line. According to Henderson et al. [1] observers will acquire objects properties within a radius of 4° of the fixation point. By fixating on the neck observers can acquire information from the horizontal stripe and the face. We also have a number of fixations to the background just left of her face and above her right shoulder, where the background is most uniform. The skintone on her torso also receives a high number of fixations. Some observers have seen on the couch in the right of the image, the same with areas in the background and the hair. The blue feathers have mostly been avoided by the observers, this could be because it is very difficult to spot a difference here.

Cartoon

The area between the face and sign gets the most fixations, containing a uniform background. The background get a high number of fixations in this image, probably mainly because this is the easiest way of seeing a difference. The same discussion as in the JP scene is valid here, where observers can acquire information from a point within a radius of 4° [1]. This enables the observers to get information about the face and sign, without moving the fixation point. Godijn and Theeuwes [83] stated that when two nearby locations are strongly activated the eyes will typically land somewhere between the two location. It is clear that the observers have been looking at both the face and sign, and therefore the center of attention will be between these regions. These are the most semantic regions and provides the most information to the observers. The tshirt get some attention, this is also one of the regions altered. This is probably fixated due to the content and the modification to the region. The shorts also receive some attention, but this is not a region altered, this only change in the image with global changes. Some observers have also looked at the hair of the character, this area is one of the larger and uniform regions.

Overall observervations

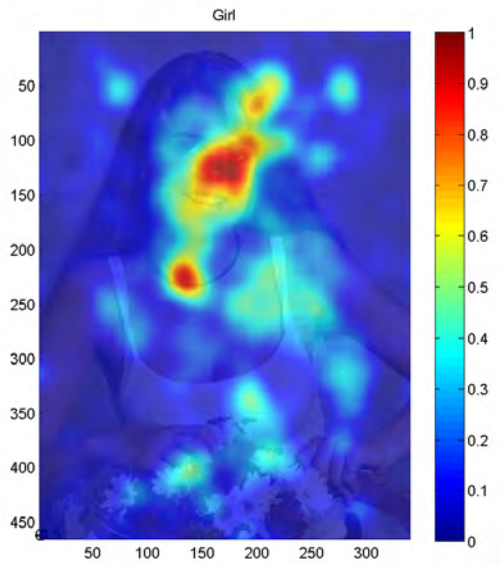
We clearly see that the face is an important region, this result has been stated in several research papers [7, 33, 84, 85]. Attention from the observers is drawn toward areas with semantic information. Attention will also be drawn toward an eccentric loci [46], an area where a visible change has been made. The results indicate that observers also acquire information within an radius of the fixation point [1]. In scenes with a uniform or almost uniform area this will be important, but not as significant as faces.

5.1.3 Eye tracker: gaze marking

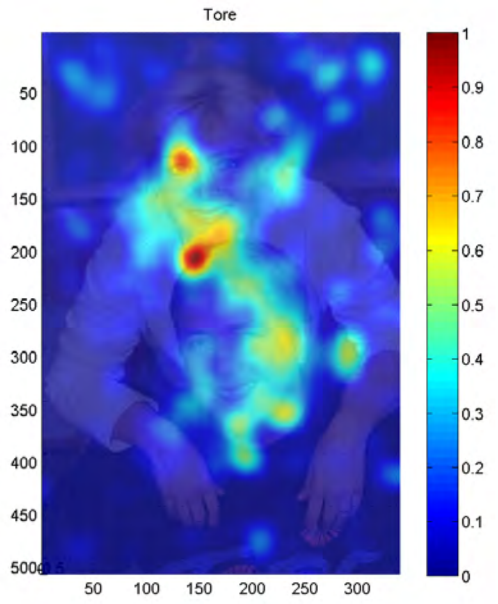
The data from the eye tracker in the gaze marking is presented in Figure 26. The observers were asked to look at the regions important for their decision in the pair comparison task.

Girl

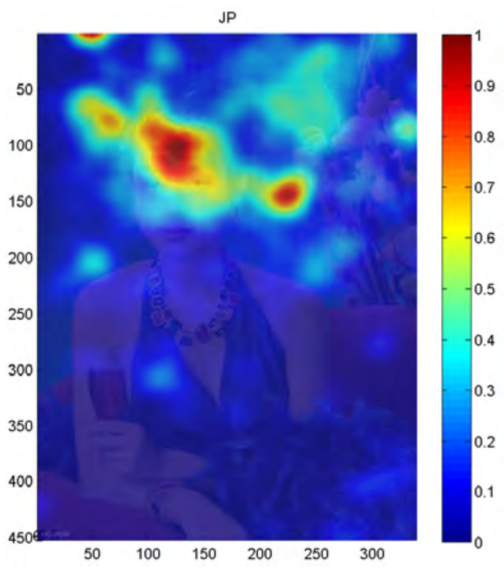
Many observers have been looking at the face in this scene, and is the center of attention. There is also a part on her neck where the necklace is located that receives many



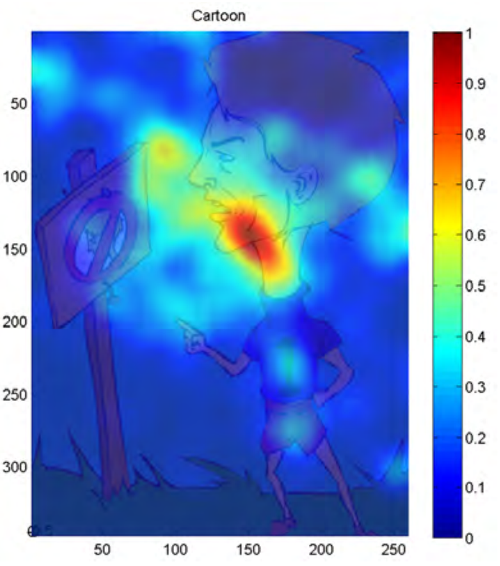
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 26: Eye tracker map from gaze marking.

fixations. A part of her hair has also been looked at by some observers. Areas like the background, chest (mostly skincolor), flowers and the white top has been fixated approximately 40% by the observers.

Tore

In the Tore scene most of the fixations are located in the faces of the girl and the guy. The center of attention is located between the two faces, just beneath the girl's chin. The reason for this could be the use of peripheral vision to obtain information about the area [1]. Henderson et al. also found out that observers will fixate between informative areas [83]. This area is located between 3 different areas, the girl's face, the guy's face and the girl's white sweater. All these areas are informative to the observers. When looking closer to the image, for the girl the left side of the face and lower part of the face get the most attention. While for the guy the right side of the face receives the most fixations. For the guy, his right side is closest to the camera and can therefore draw more attention.

JP

The forehead and region around the eyes is the center of attention in this scene. A region in the background between the head and the flowers receives a lot of attention by the observers, an area of the background up left is also looked at. The flowers in the right top corner is semi-important for the observers. We see that the observers mostly look at the upper half of the image, leaving the areas beneath her shoulder mostly unnoticed.

Cartoon

In the Cartoon scene center of attention is located in the lower end of the characters face. The background between the head and sign is also paid attention to. Semi important areas here is the tshirt with the number, the character's shorts, sign and some background areas.

Overall observations

Faces are clearly important when observers are told to gaze mark important regions. Semi-important regions are the background and some objects drawing attention, like the flowers in the JP scene and sign in the Cartoon scene.

5.1.4 Observer stated important regions

Figure 27: Example region map.

Observers marked regions important for their decision in the pair comparison experiment, this was done on paper by marking with a pen. The regions were then transformed into a binary map, with 1 (white) where the observers marked and 0 (black) for the regions left out. Figure 27 shows an example of a map from an observer. Where the observer has marked the background, face and chest area. From the 25 observers an average map has been computed, this map has then been normalized by the number of observers. The map was computed as

$$M_{observer}(x,y) = \frac{\sum_{k=1}^{N} m_k(x,y)}{N} \tag{5.2}$$

where

$$M_{observer}(x,y) \in [0,1]$$

and m_k is the individual maps for each observer and N is the total number of observers. The maps from all observers are summed and then normalized by the number of observers (N). This was done for all 4 scenes, the resulting map for all observers is shown on Figure 28.

Girl

For the Girl image the average map is shown in Figure 28(a). From this we can see that observers stated the background as important, especially on the right side, this is probably because this area is larger than on the left and visually more uniform. Facial features get a high score together with the skin tones on the chest. Observers have also marked forelock of her hair as important. Areas rated as semi important are the arms, blouse and hair. Areas like the hands, the flower bouquet and lower end of the background are not marked as important. The difference in the background is easier to spot in the high end of the image, and the flower bouquet has a lot of details without any large uniform colors making it visually difficult to see a difference.

Tore

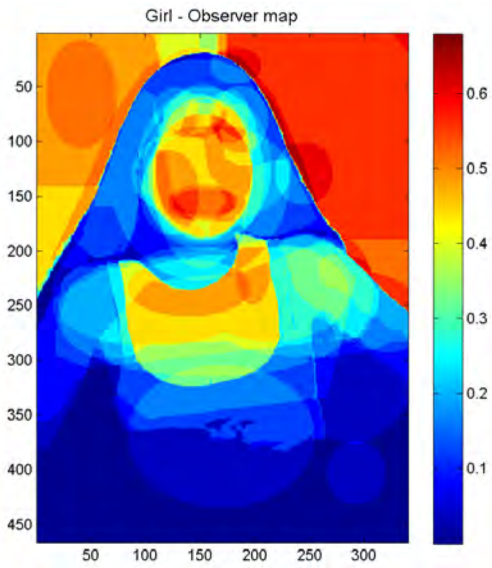
For the Tore image observers stated the faces as important as seen on Figure 28(b), both faces receive approximately the same importance by the obersers. On the guy we can see that the cheek closest to the camera gets the most attention, this is also the biggest area in the face, and for the girl observers mark more of the face. Semi important regions here are the background on the left, which is the most uniform and largest area in the background. Many observers stated the small shadow areas on the arms of the sweater as important to them. The lower end of the background has not been rated as important. The guy's sweater goes unnoticed by the observers, in this area it is difficult to tell a difference. The writing on his sweater does also go unnoticed, lightness changes in this area are difficult to notice.

JP

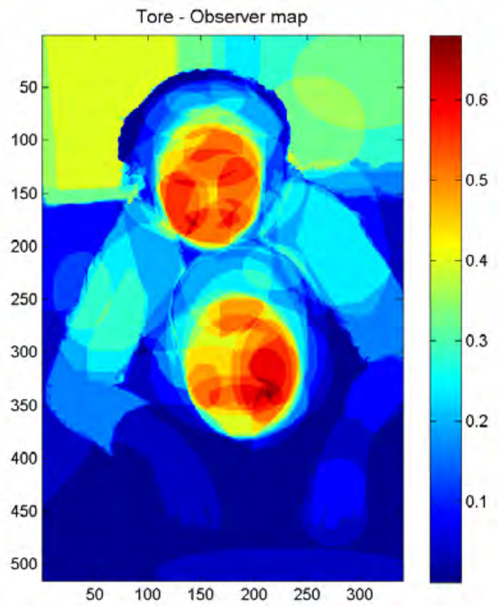
We can see from Figure 28(c) that the face is clearly important for the observers, the area around the eyes get the highest importance. The skin tones on the shoulders, especially the left shoulder get a high score. An area above the shoulder and left of the face is also marked by many of the observers. This area is fairly uniform and can therefore be easier to spot a difference in than other parts of the background. The skintone area on her neck is marked by some of the observers, this is probably because this is located within the horizontal stripe. Semi important areas are the details and the background located in the top right corner, but also the top that the girl is wearing. Observers have also marked an area on the sofa as semi important, here is both the horizontal and vertical stripe present. For the vertical stripe this area is the one easiest to spot a difference. The feather details located at the bottom of the picture seems to go unnoticed by the observers, in this area it is very difficult to see differences. The background in this image has not been rated as important by the observers, probably due to the non uniformity making it difficult to spot a difference and that other areas are easier to tell the difference in.

Cartoon

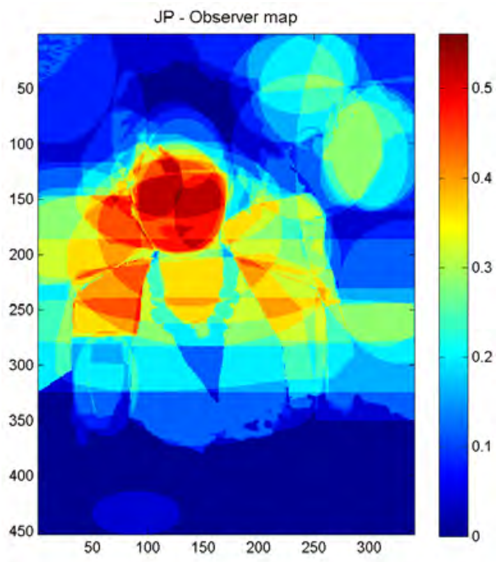
The region given the most attention by the observer in the Cartoon scene is the background in the top left corner as seen on Figure 28(d), this is probably because of the uniformity making it easy to tell if there is a difference. Another important region is the hair of the cartoon character, also a uniform area. Some of the background below the finger of the character has also been given more attention than some of the other regions.



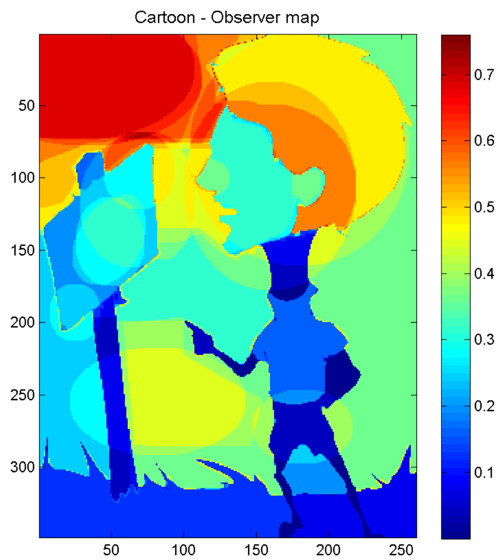
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 28: Observer stated regions.

All of these regions rated by more than 45% of the observers are large and uniform. In these areas it is easier to spot a difference than in more complex and non uniform areas. The ”stop” sign has also been noticed as important for some of the observers along with the face of the character, these two areas have more semantic value than other areas. The thsirt, one of the regions altered, has only been rated as important by approximately 15% of observers. The pole of the sign, which is a part of one of the regions altered, is not very important. The skintone of the character on the body, together with the shorts and grass goes almost unnoticed.

Overall observations

Observers tend to prioritize the background the more uniform this is. In Cartoon scene more than 70% stated the uniform background as important. In the Girl scene where we have a background close to uniform more than 50% stated this as important. The Tore scene have more details in the background and the percentage of observers stating this as important are lower than in the Girl and Cartoon scene. In the JP scene people stated the background as less important, this could be because of the ”gradient” like background with a lot of details. Another explanation could be the very visible regions altered in parts of the image, therefore drawing attention away from the background.

It is also clear that faces in the portrait images are important, but in the Cartoon image this part was not rated as important compared to the other scenes. In the portrait images the face has several elements and details that observers could find important, as shadow details and skintone color [49].

We do also see that very visible changes in small regions of the image draw attention, the horizontal stripe in the JP is stated by several observers as important even though they have problems marking the stripe exact. The skintone areas on the shoulders and upper part of the chest are stated as important, probably because this is the area where the changes are most visible.

5.2 Difference between maps

This part investigates if there are any difference between different maps. Does the eye tracking data correlate with the regions stated by the observer? Is there a difference in where different groups are looking at the images?

To evaluate the difference between the maps we use the 2-D correlation coefficient in Equation 5.3. This has been used to find correlation between fixation maps before by Babcock [33]. The equation is sensitive to position and rotational shifts and provides a first-order measure of similarity between two grayscale images [86].

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \bar{A}\right)\left(B_{mn} - \bar{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \bar{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \bar{B}\right)^2\right)}} \tag{5.3}$$

where A is one map and B is another map. $_m$ and $_n$ indicate pixel position and $\bar{A}$ and $\bar{B}$ indicate mean value in the matrix. This value will be between 0 and 1, where 0 indicate null correlation (completely different maps) between the maps and 1 indicate full correlation (identical maps).

5.2.1 Observer maps for different groups

The regions stated by the observers were used to see if there were any correlation between different groups of observers (Table 9). People who had participated in the eye tracking experiments before were not investigated due to the low number of people who have participated in eye tracking experiments before.

Observer maps	Girl	Tore	JP	Cartoon
Expert/non-expert	0.7426	0.8204	0.4521	0.6952
Psychophysical experience/no experience	0.3169	0.3902	0.5636	0.6861
Recognized Tore/did not recognize	x	0.7870	x	x
Seen image/not seen image	0.6766	0.6787	0.6466	0.2365

Table 9: Observer stated region correlation. Closer to 0 indicate different maps and closer to 1 indicate similar maps.

Expert and non-expert

People who are experts were compared to people who are non-experts. Experts are observers who have studied color science, background in photography color imaging or similar fields. Non-experts are observers with little or no experience in the field of color.

The results show that the Tore scene had the highest correlation with 0.8204. This is probably because of the non uniform background and none of the areas have a high visible change, and therefore the attention is drawn toward the faces. The JP scene had the lowest with 0.4521, the experts have marked more details than the non-experts in this scene (Figure 29, maps for other scenes and groups are found in Appendix C).

The biggest difference between those who are experts and those who are not is the level of details they mark. The experts mark smaller and more areas with more precisness than those who are non-experts. We can see this for the Girl, Tore and JP scene, but for the Cartoon scene it's not so clear. Here both groups have marked areas in similar size and shape, this could be because of the low complexity and lower detail level in this scene compared to the other scenes. We do not find any evident for one group being more concise than the other group. In some scenes the experts agree more than the non-experts, but in other scenes this is the other way around.

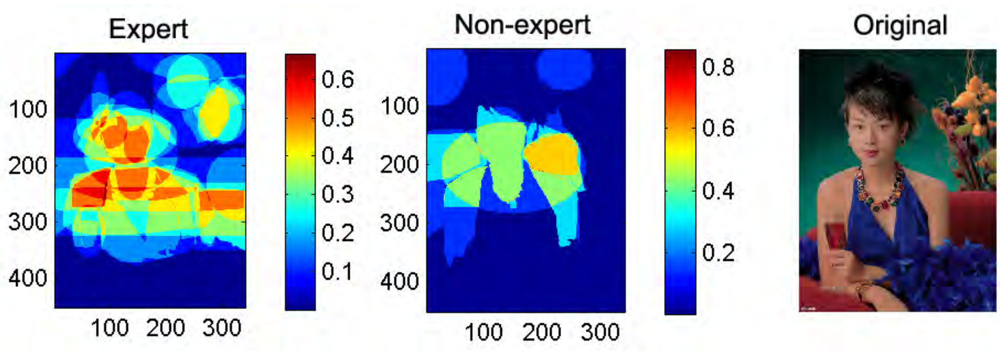


Figure 29: Observers map from experts and non-experts in JP scene. Experts mark smaller and more precise areas than non-experts.

Psychophysical experiment experience

Here people who had participated in psychophysical experiments are compared to people who are new to this. From the results in Table 9 Cartoon has the highest correlation while Girl has the lowest. For the Girl scene most of the difference is in the background (as for Tore). People who have not participated in psychophysical experiments tend to mark more of the background and less in the skintone areas (Figure 30). It seems like those who are familiar to the psychophysical experiment tend to mark smaller and more precise areas, the same as for the expert observers in Section 5.2.1. This is probably because those who have psychophysical experience also are experts. For the Cartoon scene we have a more similar spread of regions between the groups, and that's why Cartoon scores the highest.

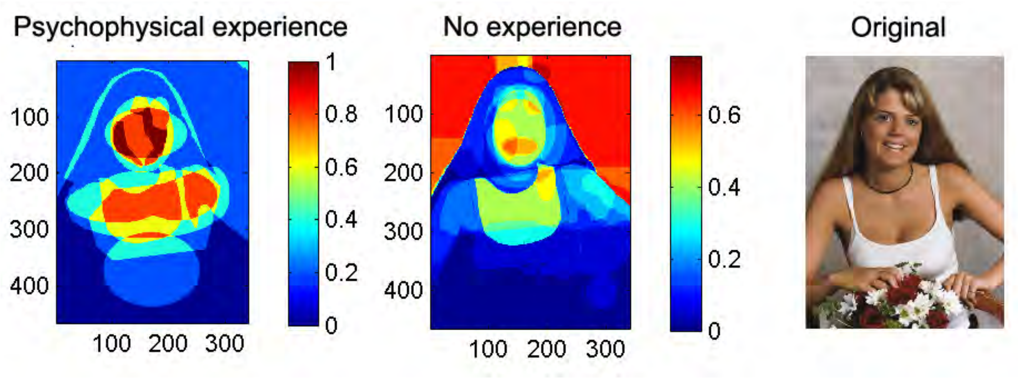


Figure 30: Observer maps from psychopysical experienced and unexperienced. Experienced mark more skintone and less background than the unexperienced.

Recognized Tore

There is a 0.7870 correlation between those who recognized the man in the Tore image and those who didn't. The faces are the most important in both groups, but some areas on the white sweater is also marked together with the upper part of the background. With a correlation this high we cannot say that a person who is recognized attract more or less attention. The biggest difference between these two groups can be found in the face. Those who recognized the man have marked more specific and precise areas than the other group. We see this in both faces, not only in the face of the man. This is then not relation with whether the observer recognized the man or not. The other scenes have not been evaluated here because this only effects the performance of this scene.

Seen images

Here we have a correlation with only minimal changes between the Girl, Tore and JP scene. This correlation indicate a moderate correlation between the groups, and no clear difference between this grouping. For the Cartoon image we only have 0.2365, but this is probably beacause only one observer had seen the image before, therefore this is not up for further discussion.

5.2.2 Eye tracking maps - binary for different groups

Here the binary eye tracking maps from the pair comparison task are compared between different groups. This map is created by using the frequency map from Section 5.1.2 and setting white pixels where the observer had been looking and black for areas with no fixations. The limit set for when an observer had been looking or not were set to 35% of the maximum value. With a threshold of 35% the area match visually across the observer stated regions and the eye tracker map. The results were a map with 1 for a pixel with value greater than 35% of the maximum and 0 for a pixel below this limit. The map was similar to the one in Figure 27. These binary maps were summarized creating an overall map for each scene. This binary method will be used later to compare observer stated regions and eye tracking maps. The results between the different groups can be seen in Table 10, as mentioned earlier 0 indicate no correlation between the maps and 1 indicate full correlation.

Binary eye tracker maps	Girl	Tore	JP	Cartoon
Experts/non-expert	0.8331	0.6183	0.5257	0.5494
Psychophysical experience/no experience	0.6784	0.5790	0.6784	0.6692
Recognized Tore/did not recognize	x	0.6776	x	x
Seen image/not seen image	0.7526	0.6158	0.6584	0.5713

Table 10: Eye tracker region correlation.

Experts and non-experts

We can see from Table 10 that the Girl scene has the highest correlation and the JP scene has the lowest. For the Girl scene the maps are very alike (Figure 31), we can see that the face area get more or less the same values. The correlation for the other scenes are lower than for the Girl image, but this could be because of the number of observers in these scenes. In the Tore scene both groups have more fixations on the top face than on the bottom face. In Cartoon and JP the number of observers who are non-experts are lower (some observers were discarded due to calibration problems, techincal problems etc.) than for the first 2 scenes, and with more observers the correlation could be higher.

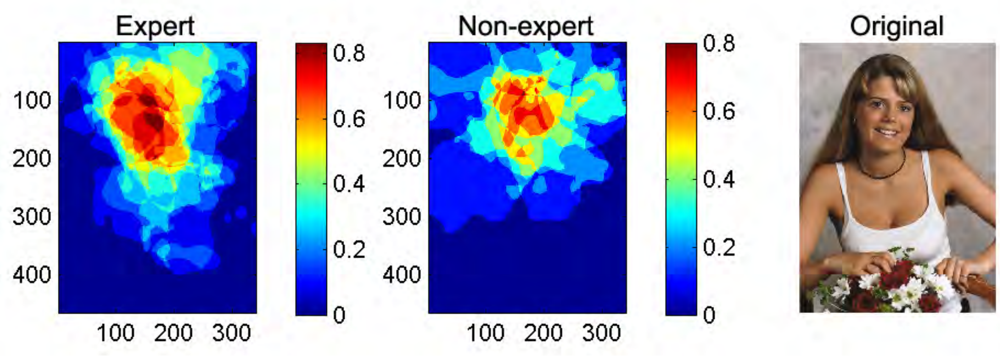


Figure 31: Eye tracker correlation between experts and non-experts in Girl scene.

Psychophysical experiment experience

The different scenes score more or less the same here, from 0.5790 to 0.6784. Here we also have the same problem as for the experts and non-experts in Section 5.2.2 with a low number of observers, but not as severe. We cannot say that there is a significant difference between those who have participated in psychophysical experiments and those who have not when it comes to the fixation maps. The same main areas are fixated between the groups, with the faces as a center of attention.

Recognized Tore

The correlation between those who recognized the man in the second scene and those who didn't is 0.6776. This indicates that there is not a large difference between these two groups. The top face is center of attention in both groups and the bottom face is the second point of attention.

Seen images

For the Girl scene we have a correlation of 0.7526, this indicate a correlation between those who have seen the image and those who have not. For the Tore and JP scene we also have a high correlation with 0.6158 and 0.6584. For all these images the same main regions have been focused. For the last scene Cartoon only one person had seen this before therefore this is not discussed.

5.2.3 Observer stated regions and binary eye tracking map

Here the regions from the observers (Section 5.1.4) are compared to the results from the eye tracker map (Section 5.1.2). These maps are build on the same principle, with a binary map where you have looked and marked or not looked and marked.

Observer and eye tracker maps	Girl	Tore	JP	Cartoon
Overall correlation observer and eye tracker	0.4784	0.5076	0.7678	0.0609
Experts	0.5114	0.5051	0.7301	0.0538
Non-experts	0.4102	0.3617	0.4276	0.1415
Psychophysical experience	0.7462	0.4936	0.6431	0.0116
No psychophysical experience	0.3826	0.3764	0.6389	0.1041
Recognized Tore	x	0.4262	x	x
Did not recognize Tore	x	0.4362	x	x

Table 11: Eye tracker and observer region correlation.

Overall correlation

The overall correlation between the eye tracking map and the observer map for the different scenes is found in Table 11. For the Girl image we have a correlation of 0.4784, and we can see from Figure 32 that the face is fixated the most for the eye tracker. This is also important states the observers, but the background is also stated as important by the observer. The background difference is the main reason why this scene doesn't score higher, but there is still a correlation.

For the Tore scene the correlation is a bit higher than for the Girl scene with 0.5076. Observers stated the 2 faces in this scene to be important, along with a bit of the background. The eye tracker map shows that only the top face (girl) is important, this area has more and longer fixations than the face of the guy. There is also a little difference in the background between these maps. There is a correlation between these, mainly due to the similarity between the face of the girl.

The JP scene has the highest correlation between the eye tracker map and the observer map with 0.7678. Observers stated face and skintones on the upper body as important, and the eye tracker map shows almost the same. Both maps indicate that skintones are an important factor. There is a little difference in the right top corner (some details

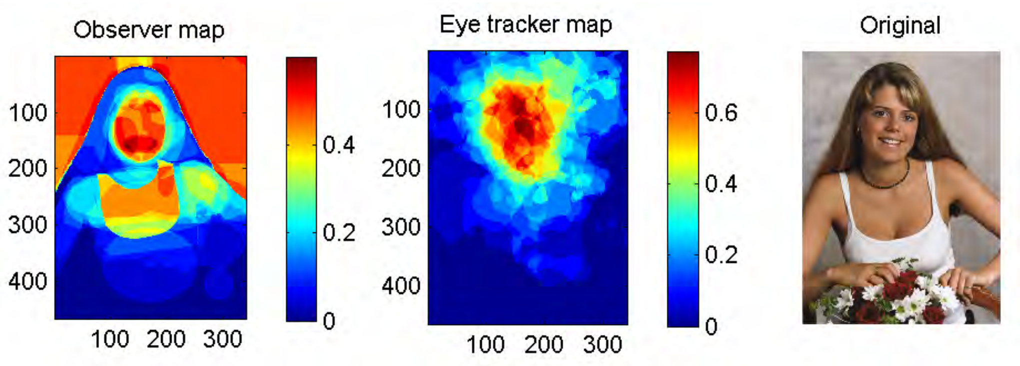


Figure 32: Overall correlation eye tracker and observer map.

in the image) and on the couch below. With these similarities we get a high correlation between these maps.

The Cartoon scene gets the lowest correlation with only 0.0609, this indicate a low correlation between what observers state as important and what they fixate on. Observers states hair and the background in the top left corner as important, but the eye tracker map shows another distribution. The background between the face and sign seems to be the area that gets the most attention. In the eye tracker map everything is centered around this point. This gives the low correlation between these two maps.

Experts and non-experts

Girl and Tore image have almost the same correlation, approximately 0.5 for the experts. For those who are non-experts the correlation was 0.4102 for Girl and 0.3617 for Tore. We see the same tendency for the JP image where expert observers get a correlation of 0.7301 while the other group get 0.4276. For the last scene Cartoon the expert observers get a correlation of 0.0538 while those who are non-experts get 0.1415. For the 3 portrait scenes there are a better correlation for the experts than for the non-experts, but for the Cartoon scene this is not the case.

Psychophysical experiment experience

Those who had participated in psychophysical experiments before had a correlation of 0.7462 while those who had not participated in psychophysical experiments scored 0.3826 in the Girl scene. For the Tore scene the numbers were 0.4936 for those with experience and 0.3764 for the non-experienced observer. For the JP scene the correlation was almost the same for the 2 groups. While for the Cartoon scene those who had not participated in psychophysical experiments scores higher than the other group.

For the two first scenes those who were familiar with the psychophysical experiments showed a correlation much higher than for those who were unfamiliar. Observers who are familiar with these kind of experiment can pay more attention to where they are looking. Similar experiments done at the school earlier have the same setup with a marking of important regions after the experiment. Therefore observers with experience can anticipate that a marking of important regions can occur afterwards. While for the JP it was almost the same. The difference in horizontal stripe is very visible and it is located in the regions where it is easy to spot a difference. This could be why the JP scene has almost the same correlation for the 2 groups. The Cartoon scene show low correlation for the overall and the experts, the same happens in this group.

Recognized Tore

Here the two different groups scored almost the same. Only a minor difference, but one group doesn't show a significant difference over the other.

5.2.4 Freeview and gaze marking

In this section we compare the freeview and the gaze marking tasks. In the freeview task the observers were told to look freely around in the images and in the gaze marking they were asked to look at the important areas for their decision in the pair comparison experiment. Score for the different scenes can be seen in Table 12.

Scene	1	2	3	4
Correlation	0.565	0.525	0.605	0.596

Table 12: Correlation between freeview and gaze marking. 0 indicate no correlation while 1 indicate full correlation.

Girl

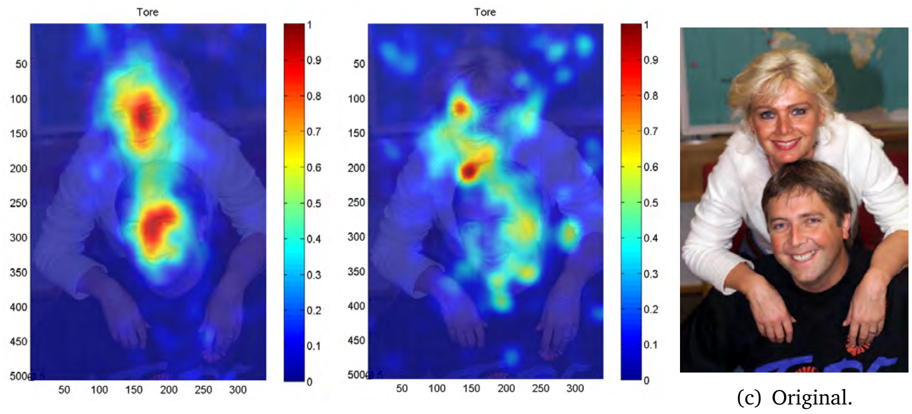
In the freeview task for the Girl scene observers mainly looked at the face. There are also some fixations on the chest skintones and on some of the flowers. In the gaze marking we do not have the same high concentration on the face, there are still a concentration here but other areas get more attention than in the freeview. There is an area around the necklace that get fixated, containing mainly skintone. There are also other areas receiving fixations that contain skincolor, like on the girl's left shoulder and down to the chest. Some fixations are also located on the girl's right shoulder, but not as much as the left shoulder. The hair does also get a fair amount of fixations, this is in the brighter area of the hair. Some of these fixations also go over into the background. In the gaze marking some observers have also been looking at the white blouse, but this was not present in the freeview. The overall correlation between these two maps were 0.565, indicating some correlation but there are also major difference in some areas.

Tore

In the Tore scene for the freeview task the observers looked at the faces of the man and the woman (Figure 33), and some minor fixations to the background. In the gaze marking we have a more spread distribution of fixations. We do not see the two main fixation areas as in the freeview task. In the gaze marking the observers have fixated in more than these two areas. The left side of girl's face and the right side of the man's face are important, but the area with the highest concentration is in the shift between the man's hair, the girl's sweater and her chin. We also have fixations on the white sweater, the background and the girl's hair. The overall correlation between these two maps was 0.525, this indicate a moderate correlation.

JP

For the freeview task on the JP scene the face got the most attention by the observers, but we do also see some fixations on the flowers on the right. In the gaze marking task the focus points are moved, the face is still important but we have a shift toward the girl's hair. The background gets more attention than in the freeview, but there are also some fixations on the flowers. Some observers have looked at the blue dress, and at the area



(a) Freeview. (b) Gaze marking. (c) Original.

Figure 33: Maps from freeview and gaze marking for Tore scene.

around her shoulders. The overall correlation was 0.605, indicating a higher correlation here than in the 2 first scenes.

Cartoon

The Cartoon scene shows a high concentration of fixations in the face of the character in the freeview. We do also see some fixations on the sign, and some on his tshirt and finger. In the gaze marking we still have a high fixation area around the face, but lower than in the freeview. The tshirt still gets some attention, and the background is fixated more than in the first task. We do also see that the sign have a lower concentration of fixations than in the freeview. This gives a correlation of 0.596 between the tasks.

Overall observations

In the freeview when the observers were told to look freely, faces was the most fixated areas. In the Cartoon scene the fixations were more spread out, but the face was still the center. In the gaze marking task faces are still important and these areas get the most attention, but areas where the observers saw changes were fixated. In a freeview without any specific task observers are consistent when looking at the images, and when they get a specific task the consistency decreases. This is the same result as Privitera et. al [46]. We do also see that in the last scene the sign gets less fixations than in the freeview, and more attention on the background. This is probably because the background is uniform and it is easier to see a difference here than on the sign.

5.2.5 Freeview and continuous eye tracker maps

Here the freeview is compared to data points from the pair comparison experiment, and the results can be seen in Table 13. The eye tracker map used here is the same as in Section 5.1.2.

Scene	1	2	3	4
Correlation	0.582	0.777	0.207	0.683

Table 13: Correlation between freeview and eye continuous tracker map.

Girl

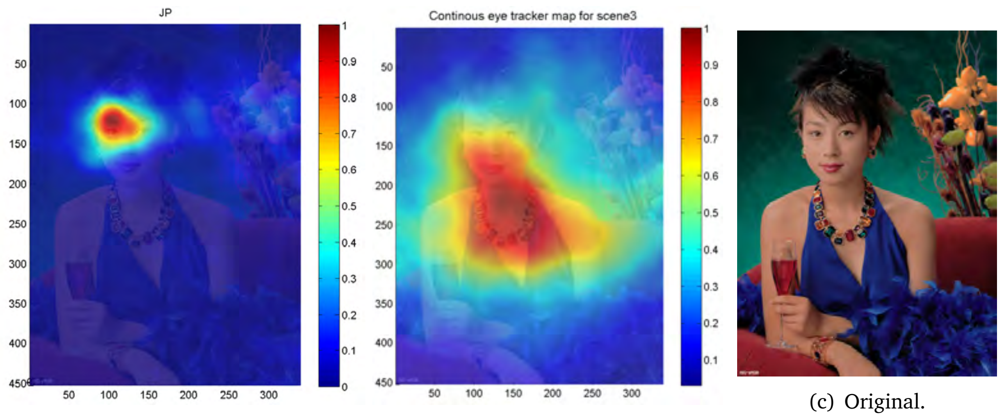
For the Girl scene both maps show the importance of the face, most of the fixations are located with the face as a center. In the pair comparison the background has been fixated more than in the freeview. In the freeview some observers also looked at the flowers, but we do not see much of trace of this in the pair comparision experiment. This is because it is difficult to spot a difference in this region, but it provides information about the scene in the freeview task. Overall there is a correlation between these maps and we get a correlation coefficient of 0.582.

Tore

In the Tore scene most of the difference is in the face of the guy. In the freeview obersvers have looked approximately the same at both of the faces, but in the map from the pair comparison the guy gets less attention and more attention is focused on the girl's face. This gives the correlation of 0.777, which indicate a medium strong correlation between these maps.

JP

The JP scene gives a fairly large difference between the two maps as seen on Figure 34. In the freeview task observers focused on the face and some at the flowers, but in the pair comparison task observers have focused on the upper part of the body where the horizontal stripe is located. This shows that introducing a highly visible modification to an image, even though the changes are small, will effect the way we look at images. This is in correlation with Privitera [46] where subjects often focus their attention on eccentric loci on the image. The difference between the maps gives a correlation of 0.207, which indicate a weak correlation.



(a) Freeview. (b) Pair comparison. (c) Original.

Figure 34: Maps from freeview and pair comparison in JP scene.

Cartoon

The Cartoon scene gives an overall correlation of 0.683, which indicate a medium correlation between the pair comparison result and the freeview task. In the freeview task the face of the character and the sign were most fixated. In the map from the pair comparison the background gets fixated the most, but the face and the sign still get a fair amount of fixations.

Overall observations

If we take a look at the overall results, faces are important in both tasks. In the pair comparison experiment observers tend to look more around in the images and more at the background. This is probably because it is easier to see the difference in the background and that observers will scan the image for areas where they spot a difference. In the JP scene where we introduced an error that was very visible with an small overall ΔE^*_{ab}. Observers fixations were drawn toward this area and away from the face. This indicate that errors in an image can change observers way of looking at images.

5.2.6 Gaze marking and eye tracker continuous

The gaze marking, where the observers were told to fixate at the regions they found important, is compared to continuous eye tracker map from Section 5.1.2.

Scene	1	2	3	4
Correlation	0.630	0.627	0.092	0.781

Table 14: Correlation between gaze marking and eye tracker map.

Girl

In the gaze marking for the Girl scene people have mainly been looking at the face, but we do also see a number of fixations in the area around her necklace and in the skintone area on the chest and shoulder. The observers have also some fixations in different areas in the background. For the pair comparison eye tracker map the face was the main focus point with smaller fixation areas to the background and chest. The correlation between these two maps is 0.630 (Table 14), that indicate a moderate correlation.

Tore

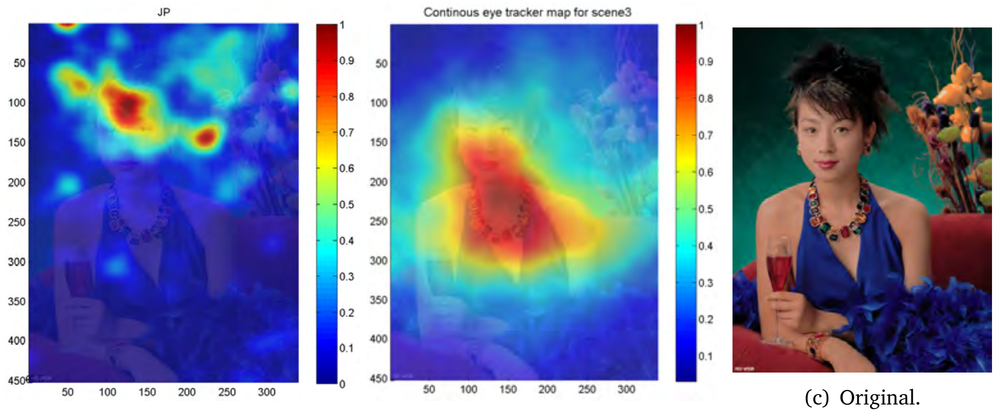
In the second scene people have still the face as a focus point for the gaze marking, but only regions in the face have been fixated on. The left side on the girl and right side of the guy is the main attraction area in the face. The area with the highest concentration of fixations is between her chin and the guy's hair. For the continuous eye tracker map the girl's face was the main focus area, with the nose and mouth as center. There are also some minor fixations to the guy's face and background on the left. This leads to a correlation of 0.627, almost the same as the Girl scene, indicating a moderate correlation.

JP

For the JP scene observers have been looking at her face and lower part of the hair in the gaze marking (Figure 35). We do also have a strong fixation area to the background at one single area. The observers have not fixated a lot at the areas where the horizontal stripe was. In the eye tracker map it's clear that observers have mainly looked at this area and in the lower part of the face. With these major differences we also get a low correlation, with only 0.092.

Cartoon

In the last scene for the gaze marking task the observers have a center point for the fixations in the neck and chin area. They have also looked at the background between the face and sign. We do also notice the fixations on the tshirt, one of the regions altered. In the continuous eye tracker map we see the same area of focus, around the neck and chin area, but we do also see a high concentration of fixations to the background between



(a) Gaze marking. (b) Pair comparison. (c) Original.

Figure 35: Maps from gaze marking and pair comparison in JP scene.

the face and sign. With these common areas the correlation is 0.781.

Overall observations

The difference in the JP scene is the most interesting. In the eye tracker map observers have been looking at the horizontal stripe. When they were told to look at the important regions, the face and background were fixated more than the area covered by the stripe. The observers notice this when they are choosing the images, but they do not find it important for their choice. Faces are also important regions in both tasks.

5.2.7 Observer regions and freeview

The maps from the observer stated important regions are compared to the freeview task. The results can be seen in Table 15.

Scene	1	2	3	4
Correlation	0.088	0.593	0.404	0.084

Table 15: Correlation between observer regions and freeview.

Girl

In the first scene observers marked the background as important together with the face, we do also see that the skintone on the chest gets a high weighting. In the freeview task observer mainly looked at the face, there are also minor fixations to her torso. With these major differences in the maps, the overall correlation is only 0.088.

Tore

For the Tore scene observers marked faces as important, together with the background as a semi-important region. In the freeview observers mainly focused on the faces. In both cases observers had faces as the two important areas, this leads to a correlation of 0.593.

JP

The face and skintone on the girl's upperbody were the regions rated as the most important by the observers after the experiment. They also marked a semi important area in the top right corner. From the freeview observers focused on the face, but they also had some fixations on the details in the top right corner. With the difference in the skintone on her upperbody the correlation is 0.404.

Cartoon

For the last scene, Cartoon, observers marked the background as the main region. Some part of the hair was also marked as important by more than 50% of the observers. In the freeview task observers looked at the face of the character, but also some fixations on the sign where we have an image of the character. We see that we have a major difference in these two maps, and the correlation is at the same level as the Girl scene with only 0.084.

Overall observations

In the freeview when the observers didn't have any spesific task, faces was the main attention area. In the marked regions after the experiment, observers marked faces as important, but also other areas like the background. In the Tore scene where we have a background without the uniformity as in Cartoon, observers both looked at and marked faces. In the Girl and Cartoon we have a more uniform background and the more uniform the background is the more attention it draws.

5.2.8 Observer regions and gaze marking

The observer marked regions and the gaze marking should give fairly similar results, the only difference is that the observers were told to mark them with a pen and in the other the observer is told to look at the important regions for his choice. Results are shown in Table 16.

Scene	1	2	3	4
Correlation	0.170	0.451	0.268	0.136

Table 16: Correlation between observer regions and gaze marking.

Girl

In the Girl scene observers have rated the face and background as important areas, the chest has also been marked as fairly important. In the gaze marking observers looked at the face and an area around the necklace. We do also see some fixations in the background. The main difference between these two maps is the background, and because of this we only get a correlation of 0.170. This indicate a big difference between marking the regions with a pen and gaze marking the same areas.

Tore

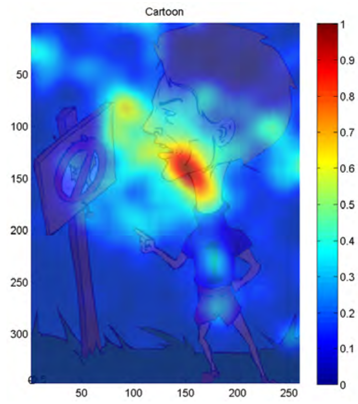
For the Tore scene observers marked the 2 faces as important, the whole girl face and mainly the right side of the guy's face. The background has been rated as moderate important. In the gaze marking observers looked at the girl's face, mainly on the left side, and the guy's right side of the face. The main focus area for the observers is still located between the two faces. We do also have minor fixations in the background. This gives a correlation coefficient of 0.451.

JP

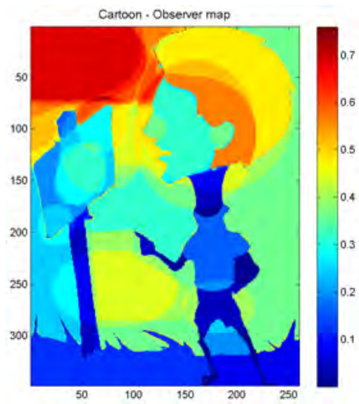
In the third scene observers marked the face and skintones of the upper body as important. They also marked an area on the top right as semi-important. In the gaze marking the face and hair is the main focus point together with an area in the background. We do also see a semi important area in the top right corner, the same area as the observers marked. The overall correlation is calculated to 0.268. The low correlation were is mainly because of the difference on the upperbody.

Cartoon

For the last scene the background is the region marked by approximately 70% of the observers. Some regions of the hair have also been rated as important. In the gaze marking some parts of the background is rated as important, but the most important area is located in the chin and neck region. These differences in the two maps leads to a correlation of 0.136. This indicate a low correlation between the two maps, this is also seen on Figure 36.



(a) Gaze marking.



(b) Observer regions.

(c) Original.

Figure 36: Map from gaze marking and observer marked regions in Cartoon scene.

Overall observations

The only scene with a good correlation here was the Tore scene, where the observer both marked and looked at the faces. The Girl scene and Cartoon both have a fairly uniform background and observers marked this as important, but they did not look specifically in this region. One explanation here could be that observers use their peripheral vision to look at the background. Overall we cannot say that there are a good correlation between the regions marked by the observers and when they got the spefic task to look at important regions. This also indicate that we cannot replace the manual marking of important regions by using gaze marking.

5.3 Time and fixation analysis

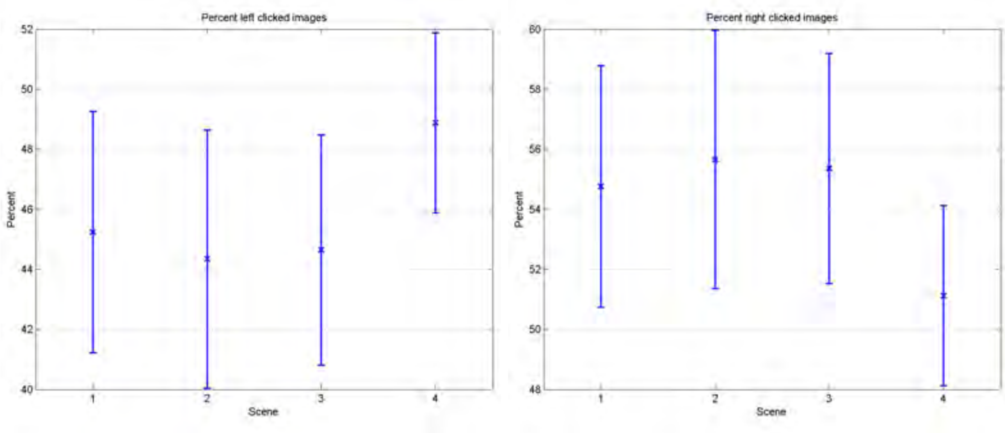
This section investigate the time used for the experiment, both in totalt and in each scene. An analysis of preferred right or left images are also done, together with a fixation analysis.

5.3.1 Left and right preferred image

During the experiment observers were told to choose the image most similar to the original, and each pair were shown twice in random order. In this section we will take a look at the distribution of the preferred left or right image. For the first scene the observers preferred the right image 54.7 % (Table 17), in the second and third scene observers preferred the right image approximately 55%. In the last scene there is a more normal distribution between left and right, where the right image was preferred 51%. In Figure 37 the percent for each scene with the 95% confidence interval based on the 24 observers are plotted. Only in the last scene, Cartoon, does the confidence interval cross the 50% line, which indicates a normal distribution. For the three other images the confidence intervals do not cross 50%, this indicate a skewness in right and left preferred images. The image pairs were shown random, and each pair was shown twice. The second time the same pair was shown this was flipped, resulting in the right image becoming the left and the other way around. This should result in a 50 % distribution on each side for the preferred images.

	Scene 1	Scene 2	Scene 3	Scene 4	Total
Left	45.2788	44.3866	44.6840	48.9219	45.8178
Right	54.7212	55.6134	55.3160	51.0781	54.1822

Table 17: Percent left and right preferred image.



(a) Left preferred.

(b) Right preferred.

Figure 37: Left and right image preferred in percent.

5.3.2 Left and right preferred for different and similar images

Some of the images in the pair comparison experiment are similar and therefore difficult to differ. 16 pairs for the Girl, Tore and Cartoon was rated as similar (LM3 and LP3,LM5 and LP5 and the different regions), and 8 pairs for the JP scene (LM3 and LP3, LM5 and LP5, StripeLM3 and StripeLP3, Stripe2LM3 and Stripe2LP3). For the similar images we get a mean of 56% right preferred images, this is about 2 percent higher than for the total scene, while for the different images the mean is a bit lower than for the total approximately 53.5% (Figure 38). This indicate a difference in the preferred images when similar images are shown and when different images is shown, but with a 95% confidence interval these will overlap and therefore we cannot say that there is a difference.

To check whether an observer has another left and right distribution for all, similar or different images, the absolute difference for each observer between the similar images

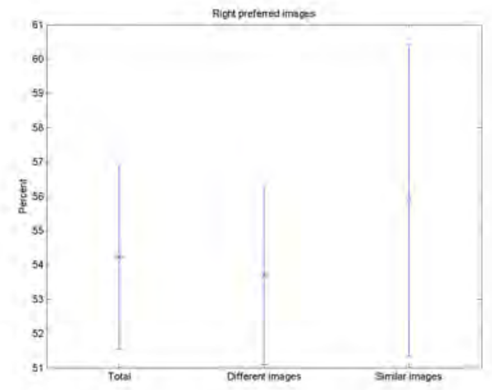


Figure 38: Percent right and left preferred image, different and similar.

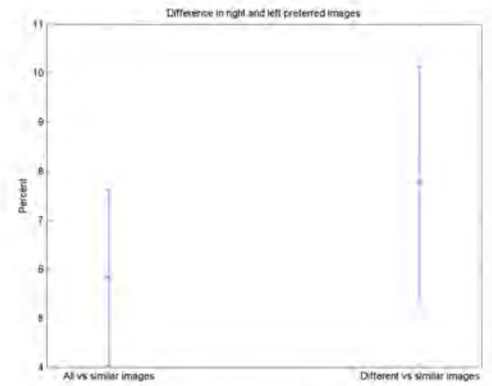


Figure 39: Difference right and left preferred image, different and similar.

and the different images have been computed. From Figure 40 we can see that observers have a difference between the similar and different from approximately 1% to 20%. This gives a mean of almost 6% difference between the similar images and the total (Figure 39). The similar images have a difference of about 8% to the different images (Figure 39). The results here indicate a difference between rating similar images that are hard to differeniate, and different images that are easier to tell apart. When the observers is forced to make a choice between two images that are difficult to differ they will prefer one of the sides more than the other.

5.3.3 Fixated left, middle and right

During the psychophyscial experiment the oberservers were shown 3 images at once, 2 reproductions on each side and an original in the middle. We have extracted the fixation points inside each of these regions. Based on this the percentage for each image have been calculated (Figure 41). For observers where we had a clear error or skewness of the fixation points due to calibration problems etc. have been discarded. The number of fixations within the image have been counted and divided by the totalt number of fixations for the 3 images. All points outside the image border have been discarded.

The overall values for the left, middle and right images are 32.4%, 34.8% and 32.8%

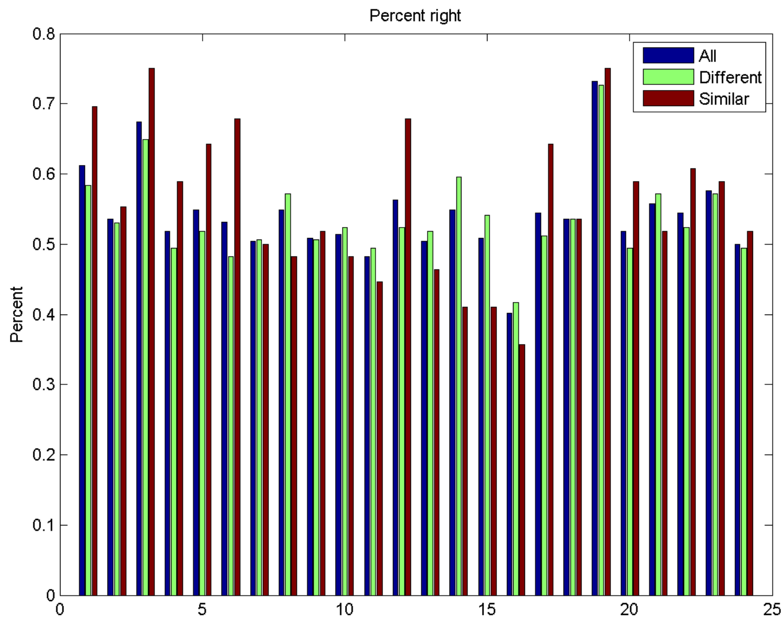


Figure 40: Difference for total, different and similar right preferred for all observers.

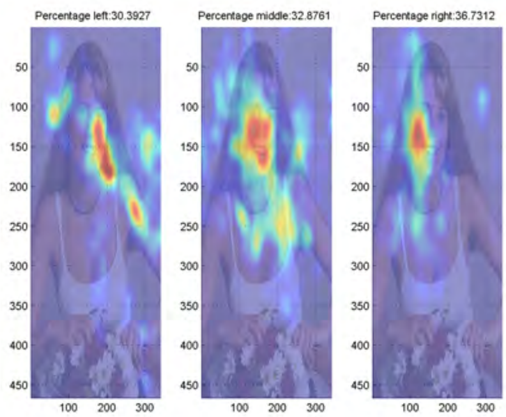


Figure 41: Heatmap for 1 observer - all 3 images.

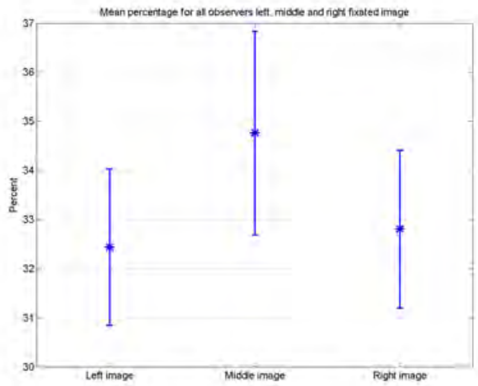


Figure 42: Mean percent fixations on left, middle and right for all images.

(Figure 42). The middle image has a higher number of fixations, and the two reproductions have almost the same number of fixations. With a 95% confidence interval there is no indications that one of the three images have more fixations than the others.

Among the different scenes the most interesting ones are the JP and Cartoon scene, where the middle image get 31% in JP and 38.5% in the Cartoon scene. In the JP scene we have a local region with a small pixel-by-pixel difference but very visible, and the observers have more fixations on each side than on the middle (original) image. While in the Cartoon scene, where the regions are small and difficult to notice, the observers have more fixations in the middle image. The 95% confidence intervals applied to the values indicate that we cannot differeniate between left, middle and right image in either of the scenes.

Fixated left, middle and right on different images

Figure 43 shows the percentage of fixations on the left, middle and right image for the images rated as different. The middle image gets the most fixations with a mean of approximately 36.4%, while the left gets 31.4% and the right 32.1%. The middle image have a significant higher number of fixations, the values here will due to the 95% confidence interval be within 34% and 39%. The left and right image cannot be differeniated. The observers use more time on the middle image than on the two reproductions. This could be because they compare each image on the sides to the original in the middle.

For the first three scenes (Girl, Tore and JP) the three images (left, middle and right) cannot be differeniated with a 95% confidence interval, but in the last scene the observers spend more than 46% of their fixations on the middle images and only 26% on the left and 28% on the right (Figure 44). One explantion of this could be the size of the image, the Cartoon scene is smaller than the other images and the observer can get an impression of the side images just by looking in the middle image and using his peripheral vision. The Cartoon scene is also a less complex image than the others, many stated the background as important here and by looking in the middle you also get an impression of the background in the reproductions.

Fixated left, middle and right on similar images

For the images rated as similar the middle images have the highest percentage of fixations, but with the 95% confidence interval it cannot be differeniated from the left and

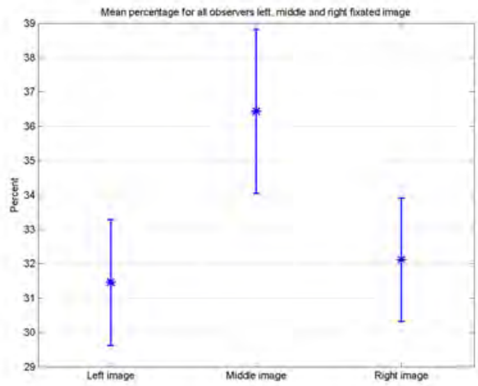


Figure 43: Percent fixations on left, middle and right for different images.

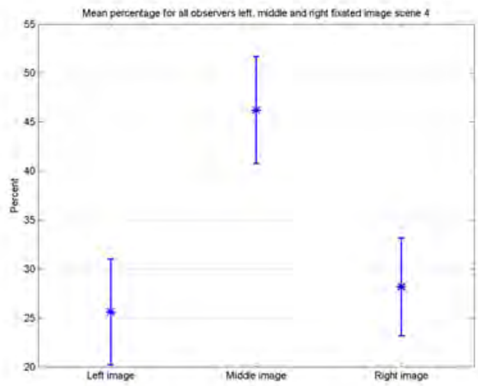


Figure 44: Percent fixations on left, middle and right different image in scene 4.

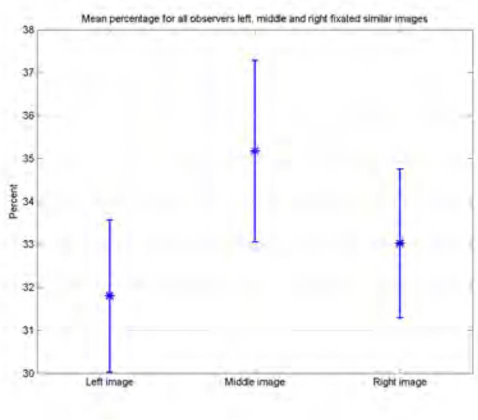


Figure 45: Percent fixations on left, middle and right similar images.

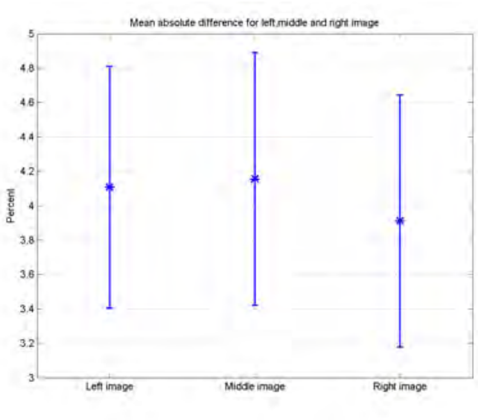


Figure 46: Mean difference fixations between similar and different images.

right image (Figure 45). The percentage of fixations on the right image has increased with approximately 1% from the different images to those who are similar. The middle image has almost the same tendency and the left side has decreased with about 0.5%. For the three first scenes left, middle and right images cannot be differeniated for the similar images, in the Cartoon scene the middle image has a mean of 39% but with a 95% confidence interval this will overlap with the left image with a mean just above 30%. The right image has a mean a little bit higher than the left but a smaller confidence interval and can be differeniated from the middle one.

Absolute difference between similar and different images

Figure 46 shows the absolute difference between the fixations in each image (left, middle and right) in the similar and different images. Left, middle and right have approximately the same difference, about 4%. Results here indicate a difference between the similar and the different images from 3-5%.

Overall observations

Observers fixate significantly more on the middle image when they have different images on each side than if they are shown 2 similar reproductions on each side. Overall results show that observers use equal amount of fixations on either side in a pair comparision experiment, and the mean of the middle image is higher than for the two sides but not necessarily a significant difference.

5.3.4 Fixation transition

The number of transitions between the left, middle and right image have been calculated for the observers with valid data points. The number of transitions have been divided by the total number of transitions for normalizing and multiplied with 100. There is some skewness in the results due to blinks, loss of eye, fixations outside the region of the images etc, but the problems are for all regions (left,right and center) and the results will indicate a correct distribution of the transitions.

Girl

In the Girl scene to the center from right or left side are almost the same, around 20 (Figure 47 and Table 18). The same from the center to the right og left side. While from left or right to right or left have approximately the same value, 10. This indicate that the observers have a transition with the middle image, and do not directly change their fixation from one side to the other side. This also indicate that the observers use the original image in the middle actively, and this provides a value for the observers throughout the whole scene.

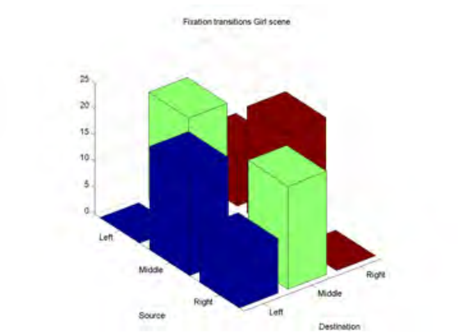


Figure 47: Fixation transitions Girl scene.

		Destination		
		Left	Center	Right
Source	Left	0.000	20.268	9.860
	Center	19.644	0.000	20.177
	Right	10.400	19.652	0.000

Table 18: Transition matrix for Girl scene.

Tore

In the Tore scene the middle image has a higher number of transitions than between the side images (Figure 48 and Table 19). From the center to the right has the highest number of transitions with 25.87, inidcating that the observers have changed their fixations from the center image to the right image a higher number of times than any other combinations. The observers also preferred the right image more in this scene than in the other scenes (Figure 37). They preferred the right image almost 56% of the times. In percent fixated the right image get almost 4% more fixations than the left image, therefore it is also normal that this image has the most transitions.

JP

For the JP scene the transitions involving the center image has values from 19 to 20.1 (Figure 49 and Table 20), while the left to right and right to left has 10.9 and 10.7. This indicate that the observer does not fixate on the left or right and change fixation point to the opposite side, but stops in the middle before fixating on the opposite or the same image.

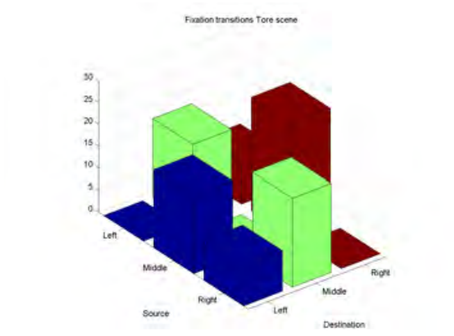


Figure 48: Fixation transitions Tore scene.

		Destination		
		Left	Center	Right
Source	Left	0.000	17.815	8.913
	Center	17.773	0.000	25.870
	Right	9.233	20.397	0.000

Table 19: Transition matrix for Tore scene.

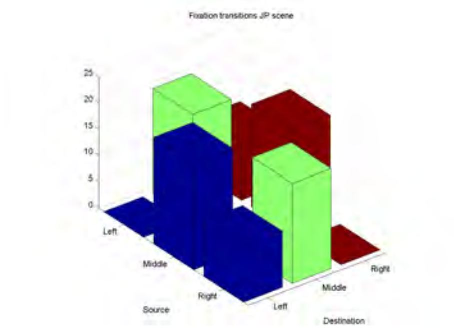


Figure 49: Fixation transitions JP scene.

		Destination		
		Left	Center	Right
Source	Left	0.000	19.776	10.929
	Center	20.181	0.000	19.420
	Right	10.679	19.015	0.000

Table 20: Transition matrix for JP scene.

Cartoon

Also for this scene we have the same distribution of the transitions as the previous scenes. All transitions involving the center image have a value around 20 (Figure 50 and Table 21). The left to right and opposite have the lowest distribution of the four scenes.

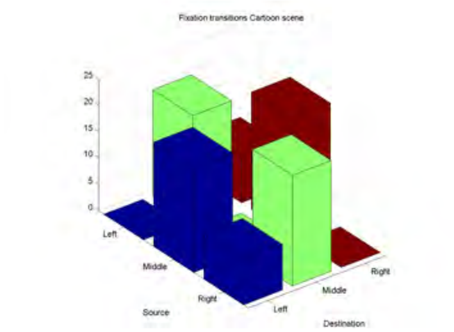


Figure 50: Fixation transitions Cartoon Scene.

		Destination		
		Left	Center	Right
Source	Left	0.000	20.000	8.408
	Center	19.748	0.000	22.194
	Right	8.660	20.990	0.000

Table 21: Fixation transitions for Cartoon scene.

Overall observations

We can see that the observers have approximately the same transitions from center to the side images and side images to center, and they do not change fixation from one side image to another as often as center to side and side to center. From these results it is clearly that the observers use the center image actively, and direct reproduction to reproduction comparison is not as important as original to reproduction comparison.

5.3.5 Time analysis

The time each observer has used in the different scenes and on the different images have been extracted (Figure 51). From the mean time used in each scene we can see that observers use the most time in the first scene, with a mean of almost 600 seconds. In the second and third scene the time is almost the same, approximately 400 seconds, and for the last scene the time decrease to a bit over 300 seconds.

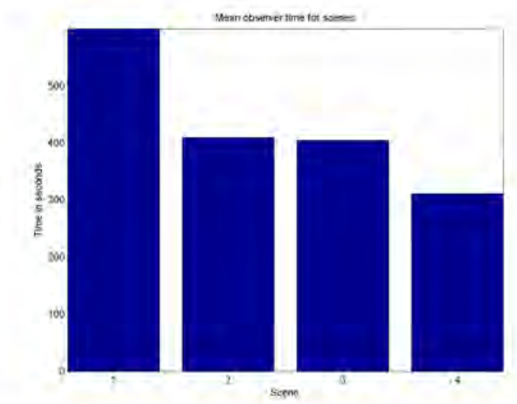


Figure 51: Mean time for each scene.

It's not surprising that observers use more time on the first scene, they adapt and learn the process of the experiment. The second and third scene is approximately the same size and complexity, while the last scene is smaller and has lower complexity. This could be the reason why obersers use less time on the last scene than on the two previous scenes.

Differences between experts and non-experts

We have analyzed the time for each scene and total for those who are experts and non-experts. In all scenes the expert observers have a higher mean than the non-experts. With a 95% confidence interval we cannot differeniate between the two groups in the first, second and fourth scene (Figures found in Appendix C.1.9), but in the third scene we see a clear difference in time used between the two groups (Figure 52). The experts have a mean just above 470 seconds, while the non-experts have a mean a little under 300 seconds. With this difference we can say that the expert observers spend more time than those who are non-experts. The reason for this could be the modification to the images, the regions altered are small and one is very visible. Radach et al. [87] found that observers spend more time on the higher the complexity of advertisements, for an expert the modification to this image could be higher than for a non-expert. From Figure 29 we can see that experts rated the horizontal stripe more important than the non-experts, but they also marked areas where the vertical stripe is located. The experts also mark more precise and smaller areas, and could be why this group spend more time then non-experts.

For the overall time the experts observers have a higher mean than those who are non-experts. With a 95% confidence intervall we can also differeniate between the two groups, the experts use significantly more time than non-experts. From Figure 53 we

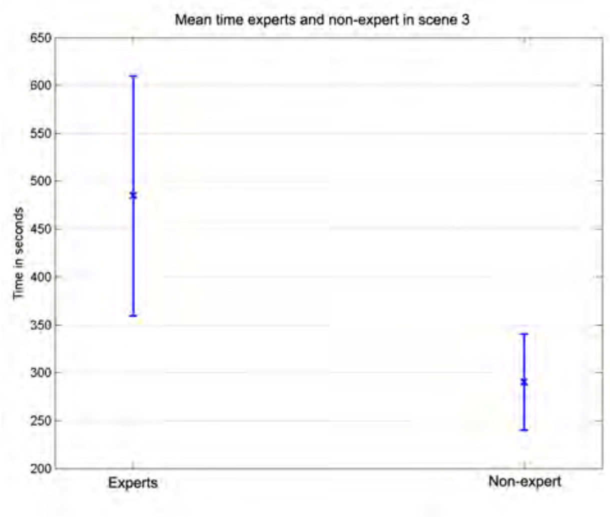


Figure 52: Time difference between experts and non-experts for scene 3.

can also see that the confidence interval for the experts observers is wider than for the non-experts, this implies more variations among the expert observers. The non-expert observers do not have the same variations in the total time used.

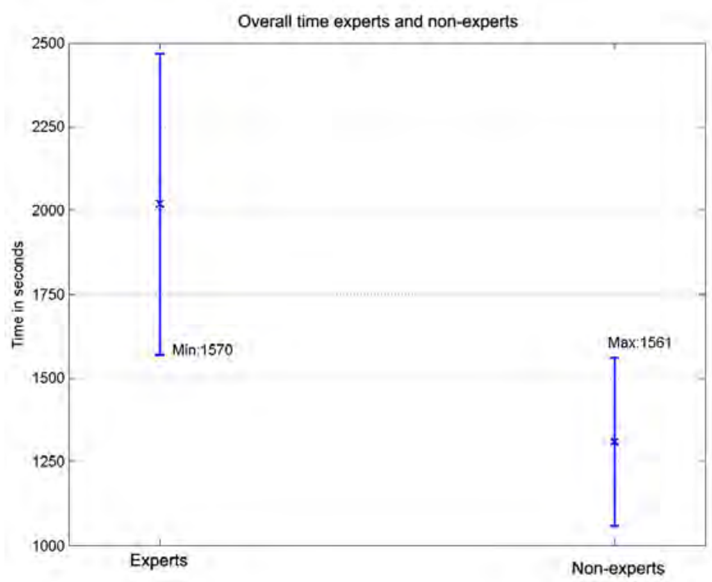


Figure 53: Total time difference between experts and non-experts.

6 Image difference metrics

This chapter implements and use different region-of-interest to improve image difference metrics.

6.1 Image difference metrics results

The image difference metrics used are ΔE^*_{ab}, S-CIELAB, iCAM, SSIM and the hue angle algorithm. The first 4 metrics are the most common and most used, therefore these have been choosen. The hue angle metric has been choosen because of its relationsship with ΔE^*_{ab} and the lack of earlier testing of this algorithm [14]. In this section these metrics were applied to the entire image, without any region-of-interest weighting.

6.1.1 Evaluating the image difference metrics

We have chosen two ways of evaluating the image difference metrics used. The squared Pearson product-moment correlation coefficient [88].

$$r^2 = \left(\frac{\sum Z_X Z_Y}{N} \right)^2 \tag{6.1}$$

This way of evaluating image difference has been used earlier by Morovic and Sun [27]. It calculates how well the data points correlate with the linear regression line. The optimal goal for image difference metrics is to match the perceived image difference linearly. The output of this function will be a number between 0 and 1, where 1 indicate full correlation and the ultimate goal for the metrics, while 0 indicate no correlation.

We have also used the Euclidean distance from the regression line. This is to check if the correlation is as a result of the points following the regression line. A high mean where indicate that the correlation is mainly due to a stochastic distribution.

$$D_{distance} = \sqrt{(}(p_i - r_i)^2) \tag{6.2}$$

where p_i is the data point and r_i is the corresponding point on the regression line.

Scene/Algorithm	S-CIELAB	iCAM	DeltaEab	SSIM	Hue angle
1	0.9413	0.6282	0.9301	0.6096	0.9634
2	0.9613	0.5683	0.9554	0.4978	0.9702
3	0.0892	0.0067	0.0794	0.0512	0.1636
4	0.9663	0.7433	0.9527	0.8467	0.9724

Table 22: r^2 value for the different algorithms. 0 indicate a low correlation while 1 indicate full correlation.

6.1.2 S-CIELAB

In S-CIELAB the lower the value the more the image is similar to the original. As we can see from Figure 54 the results are fairly linear, and the Pearson r^2 is 0.6105 (results for the individual scenes and other metrics is found in Appendix D). There are some outliers, the most obvious are the JP stripe LM3 and JP stripe LP3. These score low in the z-score, but also low on the S-CIELAB score, and are some of the best images according to the rank of S-CIELAB. All the images changed plus or minus 5 ΔE^*_{ab} are rated on the higher end of the scale in S-CIELAB, and the images with smaller regions are in the other end of the scale as expected.

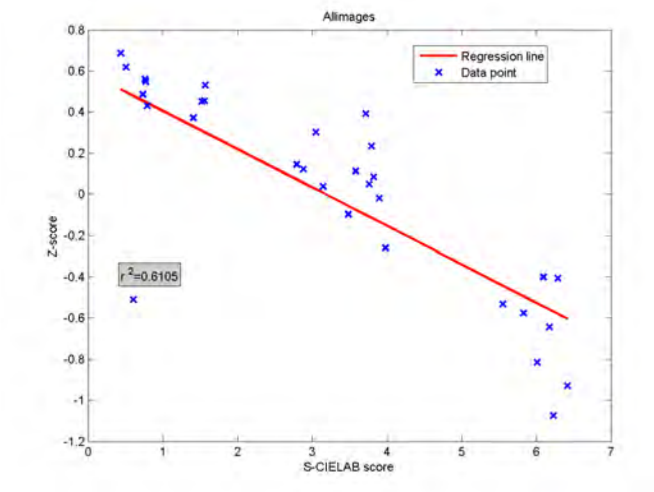


Figure 54: Z-score plotted against S-CIELAB results for all scenes.

Girl

As we can see from Figure 55 the two best images from the psychophysical experiment are also rated best by S-CIELAB, the opposite with the other end of the scale. With a r^2 score of 0.941 as seen in Table 22 we have a high degree of correlation.

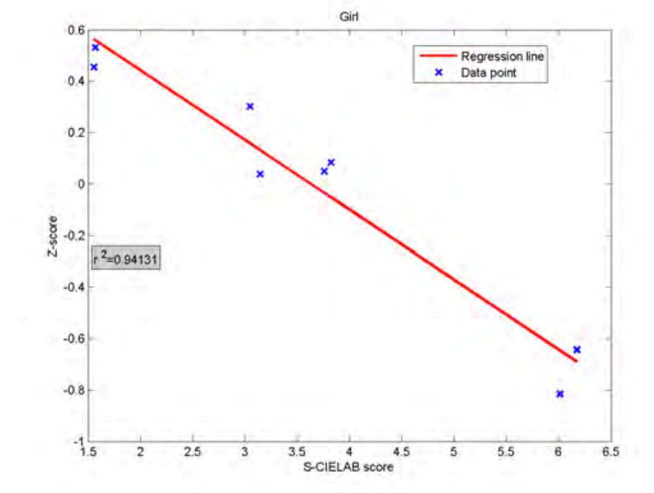


Figure 55: S-CIELAB results for Girl scene.

Tore

In this scene we have more or less the same distribution as the Girl image. The images are located more or less on a linear line, and the r^2 value indicate a high degree of correlation with 0.961.

JP

The two best images from the psychophysical experiment are rated almost identical. While the two images stripe LM3 and stripe LP3 are also rated almost identical in S-CIELAB and also better than the two previous, but according to the z-score these are among the worst images in the scene. LM3 and LP3 get almost the same value, this is also the case for LM5 and LP5. The problems S-CIELAB has here is also reflected to the r^2 value of 0.089.

Cartoon

For this scene we have most of the data points close to the regression line, and this gives a r^2 of 0.966 indicating a high degree of correlation.

6.1.3 iCAM

iCAM shows a more scattered result (Figure 56) than S-CIELAB (Figure 54) in the previous section. The lower the iCAM value the more similar the image is to the original. Some of the images with a low iCAM score have a high z-score, but other images with a low iCAM score also have a low z-score. We can see the same for iCAM regarding the two JP images with the horizontal stripe as in ΔE^*_{ab}, these score low in the psychophysical experiment and have a low algorithm score. Overall score for iCAM is 0.317, this indicate a low correlation between the perceived image difference and the calculated image difference by iCAM.

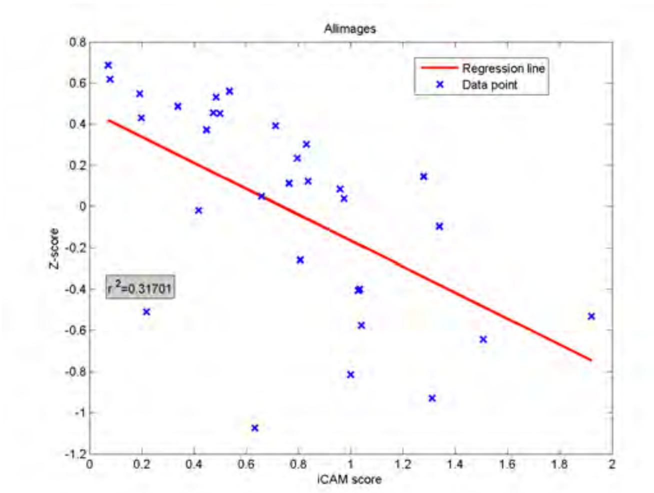


Figure 56: Z-score plotted against iCAM values for all scenes.

Girl

The LM5 image is rated almost as good as the LP3 image by iCAM, but there is significant difference in the z-score. The LP5 image is rated as the worst in iCAM, and it also receives a low z-score. Front LP3 is rated almost identical to LP3, and this correlate with the z-score. The LM3 should have almost the same score as LP3 according to the z-score, but iCAM differentiates between these. Back LM3 and back LP3 have almost the same z-scores, and almost the same iCAM result. The overall r^2 value for this scene is 0.6282, which indicates some correlation.

Tore

For the Tore scene iCAM rates LP5 as the worst image, while LM5 with the lowest z-score gets almost half the score of LP5 from iCAM and a much better rating than images significantly better in z-score. LP3 gets the second to worst iCAM score, and close to this is the front LP3. LM3 and front LM3 are given almost the same iCAM score, but in z-score they are similar to LP3 but iCAM states a big difference between these. The back LM3 and LP3 got the highest scores from the observers and the same from iCAM. These differences give iCAM a r^2 value of 0.5683.

JP

Stripe LM3 and stripe LP3 get a very low score from iCAM, but they also have a very low z-score. Their z-score is approximately the same as LM5 and LP5, but the difference between the iCAM values are great. iCAM differentiate little between LM3 and LP3, this is also in the same way as the observers. For the stripe2 images iCAM gives us a results with some difference, but the z-scores here imply that they should be close. The r^2 shows that iCAM cannot predict perceived image difference for this scene with a score of 0.0067.

Cartoon

The distribution here is almost linear except for the image with the lowest z-score. This is the LM5 image, and it is rated better than 2 other images that are better according to the observers. This gives a r^2 score of 0.7433, indicating a correlation between the perceived image difference and iCAM's predicted image difference.

6.1.4 ΔE^*_{ab}

The scores from the ΔE^*_{ab} can be seen in Figure 57. The lower the ΔE^*_{ab} value the more similar the image is to the original. ΔE^*_{ab} scores 0.626 on overall the r^2 (Figure 57), which indicates a correlation between the z-scores and the ΔE^*_{ab}.

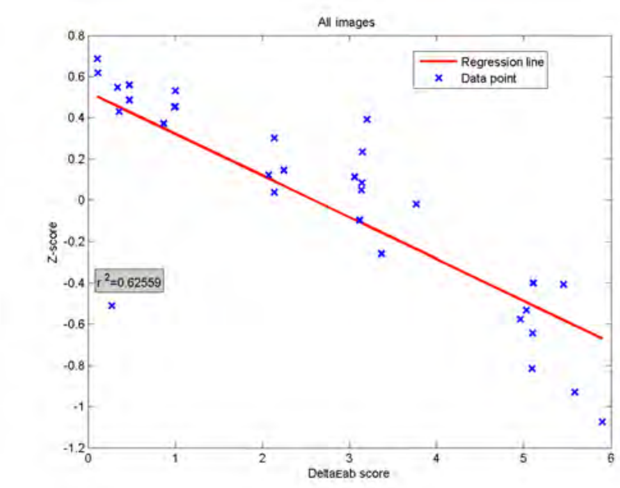


Figure 57: Z-score plotted against ΔE^*_{ab} values for all scenes.

Images with only small regions altered scores low in the ΔE^*_{ab} and have high z-scores except for 2 images from the JP scene, these two images contain the horizontal stripe. We can also see that the images altered with 5 ΔE^*_{ab} have very different z-scores. Cartoon image LM5 scores lower than -1 on the z-score and one of the JP images score approximately -0.4. This implies that the observers can differentiate between these images while the ΔE^*_{ab} difference cannot. The same is for the images altered with 3 ΔE^*_{ab}. Where we see that one image scores over 0.4 and the other scores below -0.2.

Girl

For the Girl scene the two best from the psychophysical experiment are rated approximately the same, with only a minor difference in z-score. Front LM3 is stated as the third best image by the observers and gets a ΔE^*_{ab} of 2, the same as front LP3, but this image has a z-score approximately 0.3 lower than front LM3. LM3 and LP3 have approximately the same z-score, and almost the same ΔE^*_{ab}. The same for LM5 and LP5, where the z-scores have a bigger difference than for LM3 and LP3 but not significantly different. 0.9301 as the r^2 score indicates that the ΔE^*_{ab} values correlate well with the z-scores.

Tore

The results from the Girl scene also apply to this scene. The LM5 and LP5 scores almost the same among the observers. The LM3 scores a little higher than LP3 on the z-score but not distinguishable. The front LM3 and front LP3 have almost the same ΔE^*_{ab} (some variation due to the black in this region) and score almost the same on the z-score. Back LM3 and back LP3 score best on the z-score, and the difference in the ΔE^*_{ab} is very small. This also indicate that the r^2 value should be close to 1, and with 0.9554 it gives a high correlation between the z-scores and the algorithm score.

JP

The LM5 and LP5 score almost identical on the z-scores and only minor difference in the ΔE^*_{ab} values. LM3 and LP3 have almost the same ΔE^*_{ab}. The z-scores here are different, but not distinguishable. The image rated the best by ΔE^*_{ab} is one of the images rated worst by the observers, this is the stripe LP3 (horizontal). The stripe LM3 also scores low on the ΔE^*_{ab} with only 0.45 and it has the lowest z-score in the scene. Stripe LM3 and stripe LP3 are rated almost identical ΔE^*_{ab} and only a small difference among the observers. The problems with the stripe images are reflected in the r^2 score of 0.0794.

Cartoon

This scene has a pattern similar to the other scenes, the LM5 and LP5 have a similar z-score and are located in the higher end of the ΔE^*_{ab} scale. The LP3 and LM3 have a higher z-score and these also get a lower ΔE^*_{ab} score than LM5 and LP5. The regions altered here score more or less the same on the ΔE^*_{ab} values, and only minor difference in the z-score. We also get a high score from the r^2 in this scene too, 0.9527. This is then at higher end of the scale, and indicates correlation between the z-score and the algorithm score.

6.1.5 SSIM

The SSIM algorithm is made for grayscale images. The images used in this experiment have only been altered in lightness therefore the SSIM should give a valid result. The images have been transformed to grayscale images by using the function rgb2gray in MATLAB. This function eliminate hue and saturation information while retaining luminance [89]. In SSIM a value closer to 1 indicate similarity to the original, a value closer to 0 indicate an image different from the original. From Figure 58 we can see that the

Tore scene gives very different results, all from 0.82 to 0.99. The r^2 gives a score of only 0.047, the reason for this overall low score is the difference in the results between the scenes. We can say that SSIM is scene dependent, and will not give a correct r^2 score for all results. Therefore results for each scene will be important.

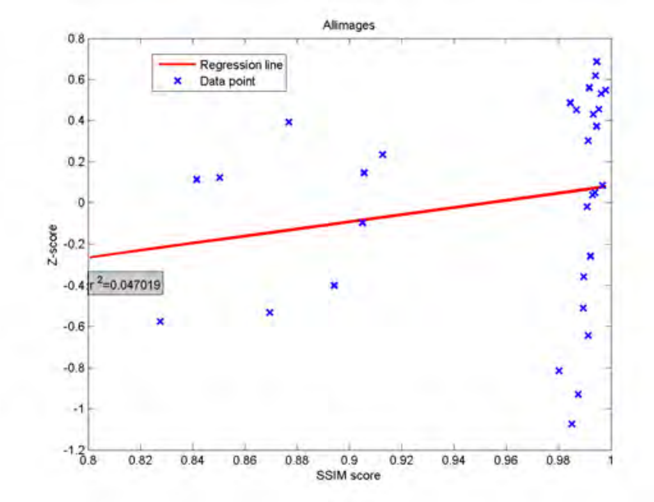


Figure 58: Z-score plotted against SSIM values for all scenes.

Girl

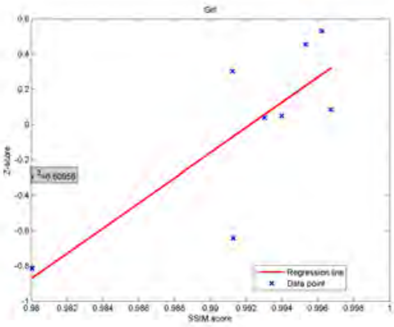
In Figure 59(a) we can see the result from the SSIM for the Girl scene. The image with the lowest z-score has been rated as the worst in SSIM, with a score of 0.9800. Next we can see that LP5 scores almost identical to front LP3, but the z-scores here are -0.64 and 0.30 which indicate that observers can differeniate between them. We do also see that the 3 images with almost identical z-score have different SSIM score and the image with the highest z-score does not score the best. If we take a look at the r^2 value for this scene we get 0.6095, indicating some correlation.

Tore

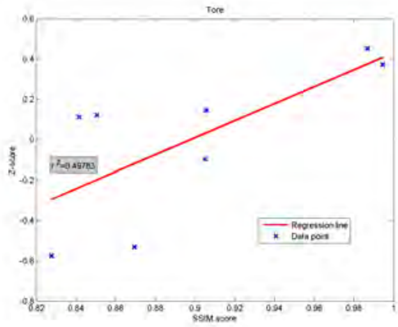
In Figure 59(b) results from the Tore scene are displayed. The image with the lowest z-score also gets the lowest score from SSIM, but the two next images in the SSIM rank have a much higher z-score than the lowest and fourth in the rank. Also for the next images, LP3 and Front LP3 score almost the same in SSIM but the z-scores are 0.1455 and -0.0972. The two last images in this scene, back LM3 and back LP3, scores 0.4524 and 0.3723 on the z-score and 0.9868 and 0.9945 for SSIM. The r^2 value of 0.4978 is lower than for the Girl image, but still it indicates some correlation.

JP

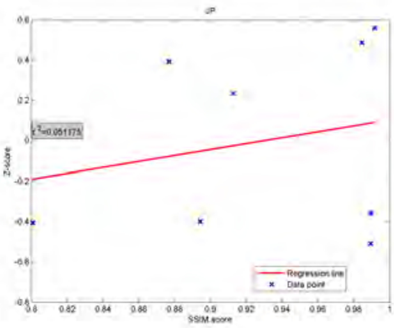
For this scene the LM5 image scored 0.8009 in the SSIM and has a z-score of -0.4059. From Figure 59(c) we can see that 3 other images have almost the same or lower z-score, but a higher SSIM score. Two of these images stripe LM3 and stripe LP3 have a z-score of -0.3575 and -0.5095 but scored almost the same in SSIM. The two images, LM3 and LP3, scored in SSIM 0.8768 and 0.9127, and the z-scores were 0.3921 and 0.2339. The image with the lowest z-score was rated as the superior one. The r^2 value indicates the same as we see in the scores, a value of 0.0511 is to low to indicate any correlation.



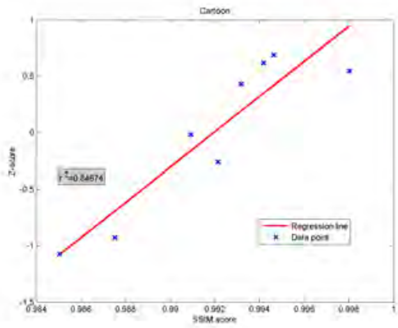
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 59: Z-score plotted against SSIM results for each scene.

Cartoon

Figure 59(d) shows the result from SSIM in the Cartoon scene. The results here are more linear than the other scenes. The two images with the lowest z-score also get the same rank by SSIM. Sign LP3 is rated as number three by the observers but gets the highest SSIM score by 0.9980. The two images with altered tshirts (tshirt LM3 and tshirt LP3) are rated fairly similar by SSIM and also by the observers. This scene gives the best result for SSIM in the r^2 with 0.846, higher than for the rest of the scenes and it indicates a strong correlation.

6.1.6 Hue angle

The hue angle algorithm gives a low score to images where only small regions have been altered, and a higher score to images where the entire image have been altered. The lower the hue angle value the more similar the image is to the original. The two images with a horizontal stripe score close to 0 in this algorithm, but the z-scores imply that these images should have a higher score from the algorithm. The hue angle algorithm differentiate well between the images with 3 and 5 ΔE^*_{ab}. All images altered with 5 ΔE^*_{ab} get a score higher than 6 while all images with 3 ΔE^*_{ab} get a score lower than 3.5 (Figure 60). This is in conjunction with the z-score from the observers. Overall correlation for this scene is 0.686, indicating a correlation between the scores.

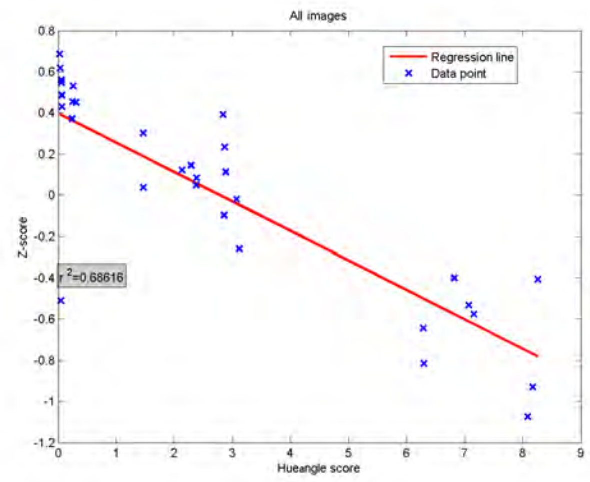


Figure 60: Z-score plotted against hue angle values for all scenes.

Girl

For this scene the hue angle algorithm gives the same score for LM5 and LP5, also the same score for LM3 and LP3. Front LP3 has approximately the same z-score as LM3 and LP3 but the hue angle algorithm differentiate between these, and gives front LP3 the same score as front LM3 even though this has a higher z-score. The two best images score almost the same and the z-scores indicate that these cannot be differeniated. The hue angle algorithm give a more or less correct estimation of perceived image difference, this is also verified by the r^2 value of 0.9634.

Tore

For this scene the LM5 and LP5 are given the lowest scores, and this correlate with the z-scores. LM3 and LP3 are given the same z-score and there is some difference in the z-scores, but with a 95% confidence interval these images cannot be differeniated (Figure 20). For the regions in this scene, regions with almost the same z-scores are also given almost the same hue angle score. The r^2 value of 0.9702 indicates a high correlation between the perceived image difference and the algorithm score.

JP

For the JP scene the hue angle algorithm differentiate between the LM5 and LP5 even though observers do not. The LM3 and LP3 are given the almost the same score, but here there are a larger difference between these than between LM5 and LP5. The 4 different stripe regions are given the same score from the hue angle algorithm, but there are a large difference in the z-scores and a r^2 score of only 0.1637. This is probably because the hue angle algorithm is based on the ΔE^*_{ab} algorithm and does not take into account the human visual system very well. Without the horizontal stripe images this scene probably would give an overall well correlation (Figure 61).

Cartoon

In this scene the hue angle algorithm does not differentiate between the images with regions, and these cannot be distinguishable by the observer. The LM3 and LP3 have some difference in the z-scores and only a minor difference in the algorithm. The same

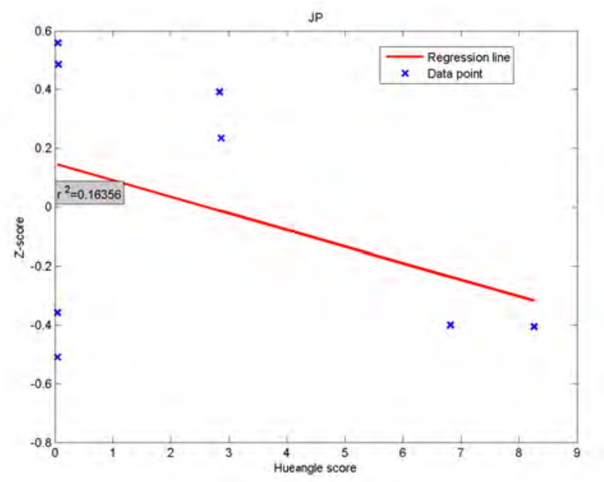


Figure 61: Hue angle scores with regression line plotted against z-score.

applies for the LM5 and LP5 images. The hue angle algorithm also scores well for this scene with a r^2 value of 0.9724.

6.1.7 Difference between the algorithms

To calculate the difference in the performance of the algorithms used, the linear regression line for each scene for the different algorithms were computed. From this regression line the mean squared difference with a 95% confidence interval were calculated as a secondary measure for performance for the algorithms (Figure 62).

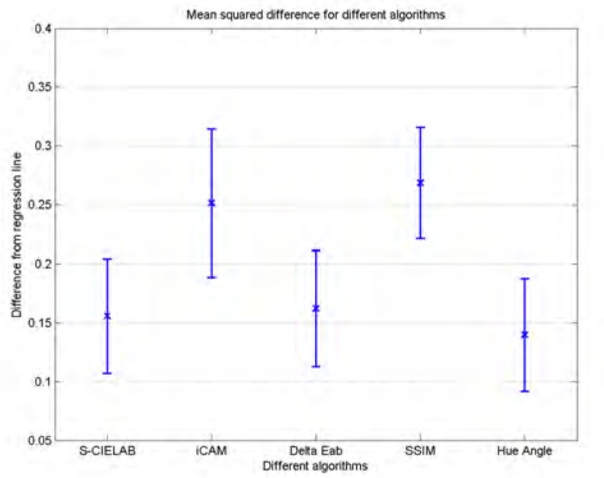


Figure 62: Mean squared difference for algorithms.

As we can see from Figure 62 the algorithms based on ΔE^*_{ab} scores better than SSIM. The hue angle algorithm has the best score, with only a little difference to the ΔE^*_{ab} and S-CIELAB. This is not surprising because of the familiarity between these algorithms, since both the S-CIELAB and hue angle are build upon the same principles as ΔE^*_{ab}. SSIM, a gray scale metric, should perform rather good due to the lightness changes, but does not come through. iCAM shows a high difference but due to the confidence interval it cannot be rated worse than ΔE^*_{ab} or S-CIELAB. For the JP image, all metrics have a high difference and the difference is especially high on the regions image (stripe LM3, stripe LP3, stripe2 LM3 and stripe2 LP3). This shows that the metrics have a problem with calculating the right score for images were small regions are altered, and the visibility of these alterations for the human eye.

The r^2 correlation indicate the same as the mean squared difference, algorithms based on ΔE^*_{ab} perform the best (Figure 63). In the 2 first scenes S-CIELAB, ΔE^*_{ab} and the hue angle algorithm outperform iCAM and SSIM. In the two last scenes there are still a difference between these but smaller than in the two first scenes.

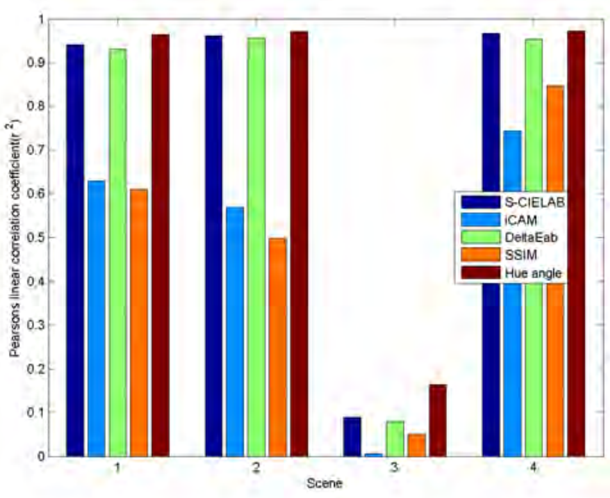


Figure 63: Whole image r^2 score for the different algorithms and scenes. The metrics based on ΔE^*_{ab} score higher than iCAM and SSIM.

S-CIELAB

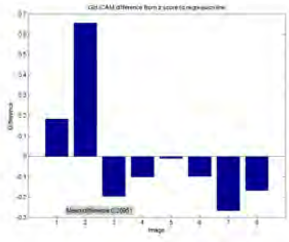
The LM3 images have a big difference in all 4 scenes, for these scenes S-CIELAB rates the images with a too low score. For the LM5 image in the Cartoon scene and Girl scene are rated as second to worst, we can also see a difference in the JP scene where LM5 gets a bigger difference than in the Girl and Cartoon scene. For the JP scene we have an overall greater difference than for the rest of the scenes, not only in the regions but also for the global changes. This tells us that for a scene like the JP this metric has problems with calculating the correct score to predict perceived image difference.

iCAM

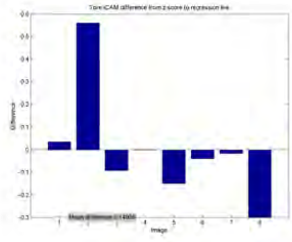
iCAM has a great difference for the LM5 images in Girl (Figure 64(a)), Tore (Figure 64(b)) and Cartoon (Figure 64(c)). In the figures the second image represent the LM5. From this score it seems that iCAM miscalculate the darkest image, and this could be the explanation for the low correlation for iCAM.

ΔE^*_{ab}

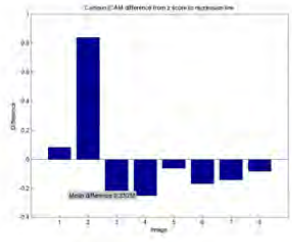
ΔE^*_{ab} has for 2 of the 4 scenes LM3 as the image with the greatest difference from the regression line. Also for the JP scene ΔE^*_{ab} has a great difference in LM3, this could imply



(a) Girl.



(b) Tore.



(c) Cartoon.

Figure 64: Difference for iCAM from the regression line. The second image is the LM5, and in 3 of the scenes this has been miscalculated by iCAM.

problems with images that are a little darker. We do not see the same difference in the LM5 even though there is a difference here too.

SSIM

SSIM was rated as worst among the metrics used. SSIM performs almost as good as the rest for the JP scene. As the other metrics SSIM has problems with the regions (Stripe and Stripe2), but it also has some problems with the global changes especially in the LM3 image and LP5 image. The same difference can also be discovered for the Tore scene, where SSIM has the biggest difference in LM3 and LP5. For the Girl scene LP5 is the image with the largest difference, and this will contribute to weaken the overall difference from the regression line. For the Cartoon scene, SSIM struggles with the small regions miscalculating the values and therefore getting a large difference from the linear regression line.

Hue angle

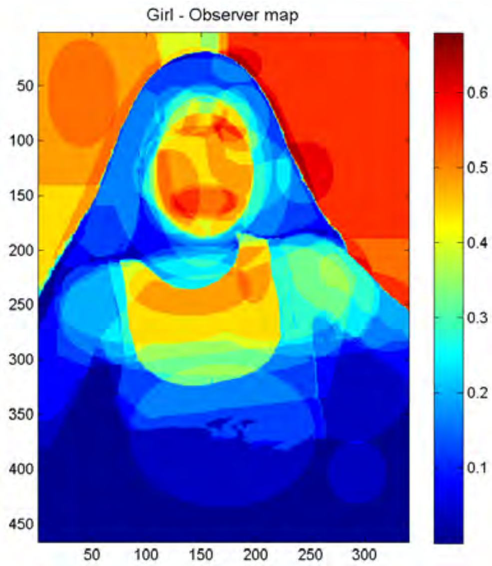
The hue angle algorithm scores overall the best among the tested metrics. It has problems with the JP scene, especially in the images with altered regions. In these images the algorithm over calculate the difference in the LM3 and LP3 images. For the Cartoon scene and Tore scene the biggest difference is in the LP3 image, and for the Girl scene front LP3 has the greatest difference. There is no clear factor saying that the hue angle algorithm has problems on some alterations on the images other than the problem with the Stripe and Stripe2 images in the JP scene.

6.2 Area based image difference

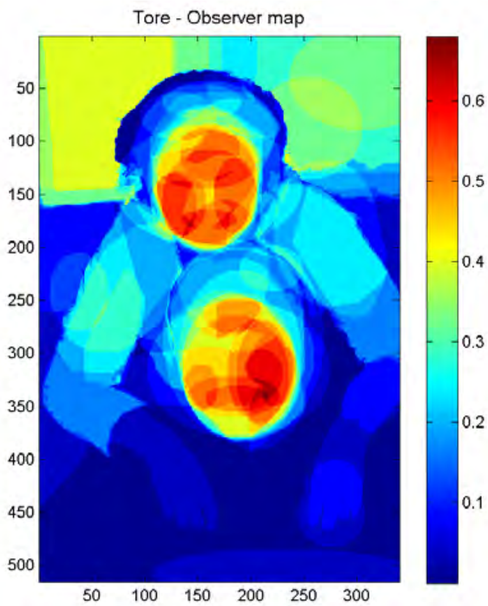
This part of the thesis applies image difference metrics to areas in the image. The areas used are from what the observers stated as important in the image and from the eye tracking data.

6.2.1 Observers stated important regions

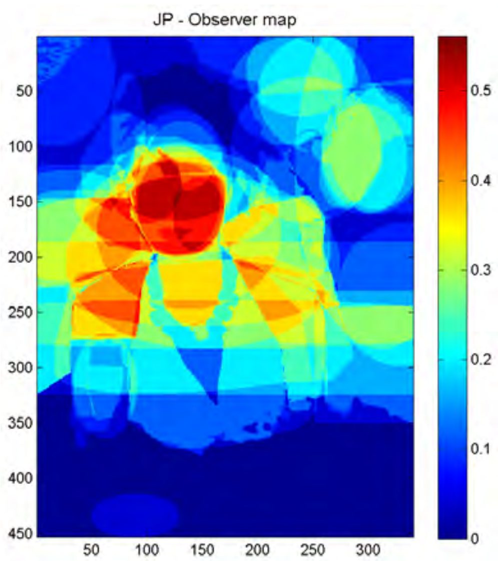
In the questionnaire given to the observers after the experiment they were asked to mark the regions important for their choice. The regions marked by the observer were used to generate a black and white image, where the white pixels are regions marked by the observer and the black are regions ignored by the observer. A region map from one of the observers is shown in Figure 66, where the observer have marked the face, background and chest area.



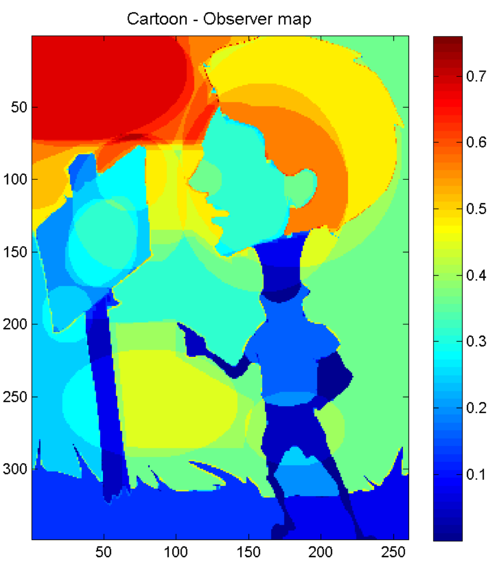
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 65: Observer stated regions.

Figure 66: Example region map.

From the 25 observers an average map has been computed, this map has then been normalized by the number of observers. The map is computed as

$$M_{observer}(x, y) = \frac{\sum_{k=1}^{N} m_k(x, y)}{N} \tag{6.3}$$

where

$$M_{observer}(x, y) \in [0, 1]$$

and m_k is the individual maps for each observer and N is the total number of observers. The maps from all observers are summed and then normalized by the number of observers (N). This is done for all 4 scenes.

Girl

For the Girl scene the average map is shown in Figure 65(a). From this we can see that observers stated the background as important, but also facial features get a high score together with the skin tones on the chest. Areas rated as semi important are the arms, blouse and hair.

Tore

For the Tore scene observers stated the faces as important as seen on Figure 65(b), these get a fairly similar score. Semi important regions here are the background and girl's sweater. Many observers stated the small shadow areas on the arms of the sweater as important to them. We can also see that the guy's sweater and the lower end of the background are not so important.

JP

We can see from Figure 65(c) that the face is clearly important for the observers, but also the skin tones on the shoulders. Semi important areas here are the details and the background located in the top right corner, but also the top that the girl is wearing, the sofa part on the right and the background just left of her right shoulder. The feather details located at the bottom of the picture seem to go unnoticed by the observers, in this area it is very difficult to see differences.

Cartoon

The region given the most attention by the observer here is the background in the top left corner as seen on Figure 65(d). Another important region seems to be the hair of the cartoon character. Some of the background below the finger of the character has also been given more attention than some of the other regions. All of these important regions are large and uniform, and it is easier to spot a difference in these areas than in more complex and non-uniform. The stop sign has also been noticed as important for some of the observers along with the face of the character.

Overall observations

Observers tend to prioritize the background the more uniform this is. In the Cartoon scene more than 70% stated the uniform background as important. In the Girl scene where we have a background close to uniform more than 50% stated this as important. The Tore scene has more details in the background and the percentage of observers stating this as important are lower than in the Girl and Cartoon scene. In the JP scene people stated the background as less important, this could be because of the gradient like background with a lot of details but also the very visible regions altered in parts of the image without the background.

It is also clear that faces in the portrait scenes are important, but in the Cartoon scene where this part was not rated as important compared to the other scenes. In the portrait scenes the face has several elements and details that observers could find important, as shadow details and skintone color, same as Bando et al. [49].

We do also see that very visible changes in small regions of the image draw attention, the horizontal stripe in the JP is stated by several observers as important even though they have problems marking the stripe exact. The skintone areas on the shoulders and upper part of the chest are stated as important, probably because this is the area where the changes are most visible.

6.2.2 Image difference metrics applied to observer stated regions

The maps created in the previous section have been applied to image difference metrics. S-CIELAB, iCAM, ΔE^*_{ab} and SSIM have been used in this section. The hue angle algorithm has not been used due to the construction of the metric, where the hue angle for the whole image is the basis. Scores for all metrics in all scenes are found in Table 23.

Scene/Algorithm	S-CIELAB	iCAM	ΔE^*_{ab}	SSIM
1	0.8932	0.5288	0.8534	0.4653
2	0.9293	0.6956	0.9201	0.8304
3	0.1409	0.0447	0.1126	0.1070
4	0.9612	0.7526	0.9475	0.6217

Table 23: r^2 value for the different algorithms, observer regions applied.

S-CIELAB

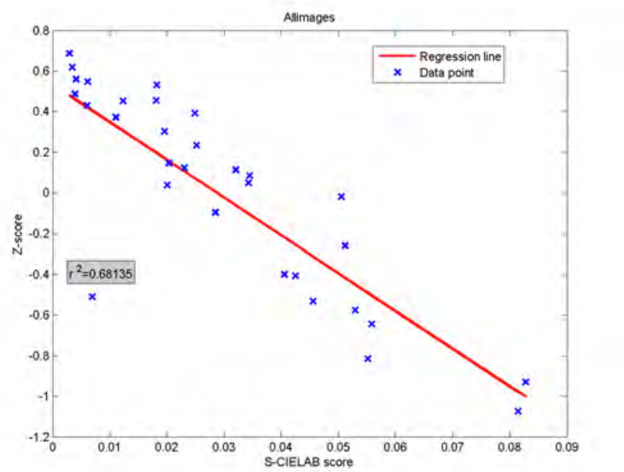


Figure 67: Observers regions: Z-score plotted against S-CIELAB.

For the Girl scene the front LP3 has been given a too low score in S-CIELAB, this score should be more like LM3 and LP3. The LM5 and LP5 have gotten approximately the same values, but in the wrong order.

The LM3 in the Tore scene receives a oto high score in S-CIELAB, this should have been more like the score of front LM3 and front LP3. The gap between the LM5 and LP5 is fairly high, but in the right order.

The Stripe LM3 and Stripe LP3 score low in the S-CIELAB but also low in the psychophysical experiment, the scores from these images should be closer to LP5 and LM5. The LM3 image should also be closer to the stripe2 images, as we can see the difference between LP3 and LM3 is small in S-CIELAB but the z-scores imply a larger difference.

For the Cartoon scene the region images score fairly similar with only minor variation and they follow the z-scores to a certain degree. The LP3 and LM3 images score almost the same, but there is a difference in the z-score here. For the LM5 and LP5, the worst image from the psychophysical experiment scores the best.

S-CIELAB scores 0.6814 in the r^2 (Table 23), we can see that most of the points are located around the linear regression line on Figure 67. Two points are located far away from the regression line, and these are the stripe images from the JP scene.

iCAM

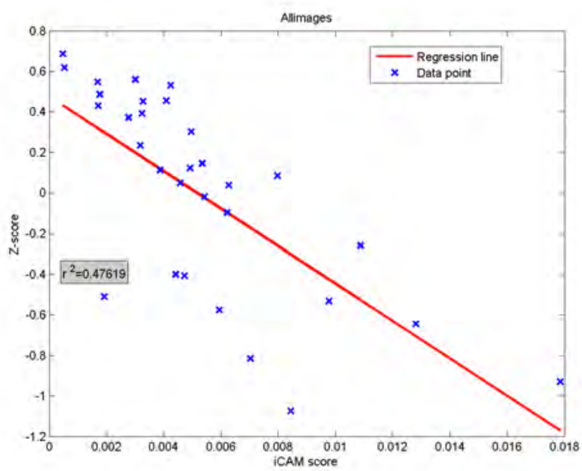


Figure 68: Observers regions: Z-score plotted against iCAM.

For the Girl scene the two best images on the z-score also score lowest in iCAM. Three images have almost the same z-score but iCAM differentiates between them. The worst image from the psychophysical experiment scores a lot lower than it should, the score for this image should have been higher than LP5 image.

For Tore the best image scores a little bit higher than the second best. The worst image, LM5 scores almost half of the score of the LP5, when they should have almost the same score according to the observers.

The best image in JP according to the observers has been rated as number 4 in iCAM. The stripe LM3 and stripe LP3 scores almost the same in iCAM, with a very low score. This score should be much closer to LM5 and LP5 according to the observers.

LM5 in the Cartoon scene gives a low iCAM score when the image gets a z-score below -1. The score for this image should be more like the score from LP5.

iCAM scores a r^2 of 0.4762, some points are located on the regression line but we see a scatter pattern (Figure 68).

ΔE^*_{ab}

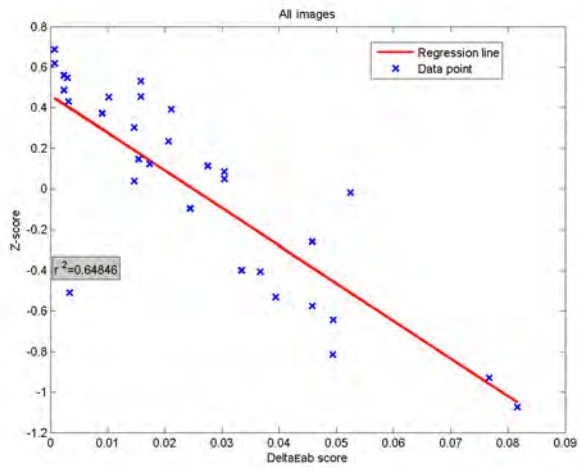


Figure 69: Observers regions: Z-score plotted against ΔE^*_{ab}.

For the Girl scene we can see that ΔE^*_{ab} rates two images with lower z-scores better than the two best. Back LM3 and back LP3 should get a score more similar to LP3 and LM3.

The LM3 image in the Tore scene seems to get a ΔE^*_{ab} score a little higher thanit should, this score should probably be more like the score of front LM3 and front LP3.

For the JP scene the stripe images should be more like the LM5 and LP5 in ΔE^*_{ab} score, and the LP3 and LM3 should probably be a little closer to stripe2 LM3 and stripe2 LP3.

For the Cartoon scene the LM3 image gets a score higher than LP3, this should have been the other way around according to the z-scores.

r^2 for ΔE^*_{ab} is 0.6485, here the JP stripe images drags the r^2 value down (Figure 69).

SSIM

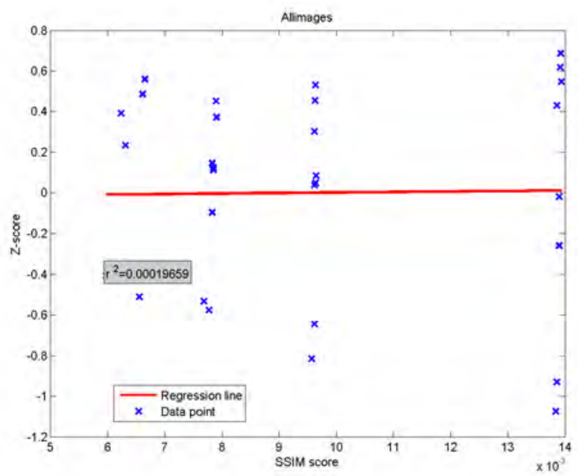


Figure 70: Observers regions: Z-score plotted against SSIM.

LM5 gets the lowest SSIM score in the Girl scene, this is correct because it is also rated as the worst image by the observers. The LP5 image gets approximately the same score as two of the images with a z-score higher than 0, the difference between these images is more than 0.6 in z-score. The image rated the highest in SSIM is not the image rated best by the observers.

For the Tore scene the two images rated almost the same by the observers get different scores from the SSIM. The front LM3 gets almost the same score as front LP3, but there are some distance in the z-scores between these.

For the JP scene the 4 worst images are spread over the entire scale. The LP3 scores second to worst from the SSIM but it is rated third by the observers.

In Cartoon SSIM rates sign LM3 to be the third worst image, while in the z-scores this score almost like the sign LP3, t-shirt LM3 and t-shirt LP3.

SSIM scores very low on the r^2, only 0.00019. The reason for this is the different scores for the different scenes (Figure 70). If we look at the r^2 value for each scene the results vary from a low correlation of 0.10 to a high correlation of 0.83.

Overall score for metrics

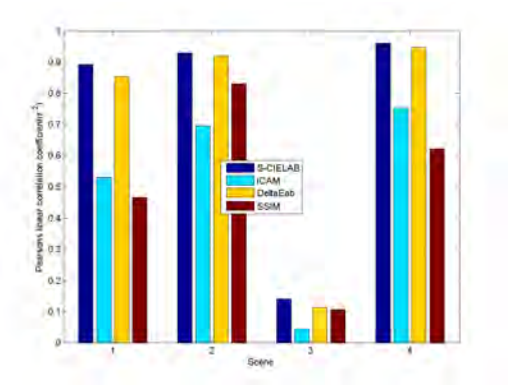


Figure 71: Overview r^2 for observer regions.

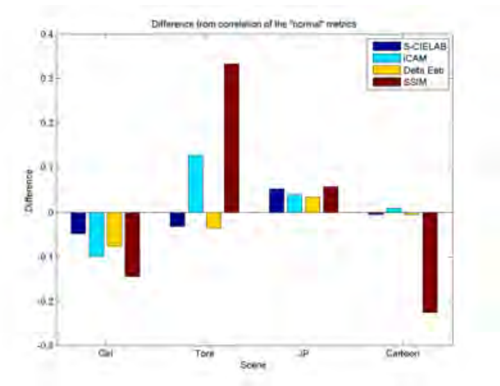


Figure 72: Difference from non-weighted map.

We can see from Figure 73 that S-CIELAB has the lowest mean, with ΔE^*_{ab} just behind. iCAM has a mean a little bit higher, but due to the confidence interval these cannot be differentiated. SSIM has the highest mean, and with the confidence interval this overlaps the S-CIELAB interval with just a bit. So we cannot state which metric performs the best when it comes to mean squared difference from the regression line.

iCAM has a high mean, this is because of the problems with predicting the perceived difference of the LM5 images. In 3 of 4 scenes it miscalculates with more than 0.45 (Figure 74), but we do not find this tendency with the other images. This miscalculation elevates the mean of the metric. Without this error iCAM would be at the same level as S-CIELAB and ΔE^*_{ab}. We do also find a tendency with the ΔE^*_{ab} too, even though this tendency is not as clear as iCAM's problem. ΔE^*_{ab} miscalculates the LM3 with 0.2 to 0.45 for all scenes, but we do not find this in the LM5 images. The reason for this miscalculation in iCAM and ΔE^*_{ab} is unclear, and more experiments must be carried out to say if this is a recurring event.

If we take a look at the different r^2 values (Figure 71), S-CIELAB has the highest score for every scene but ΔE^*_{ab} is very close. SSIM scores lower than iCAM for two of the

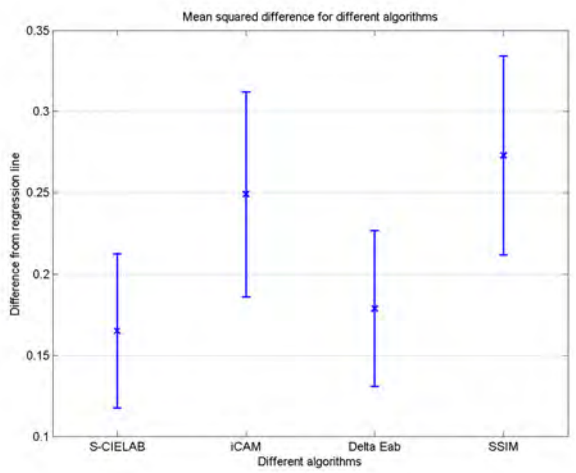
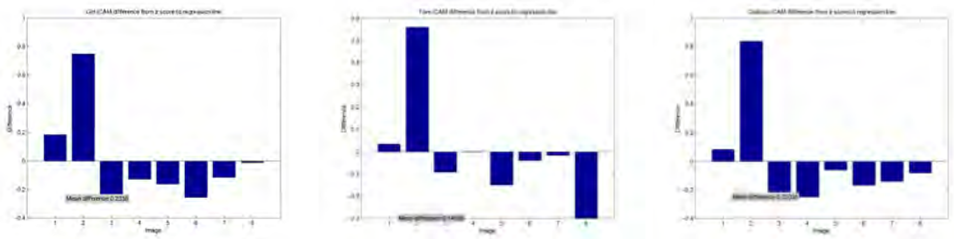


Figure 73: Mean squared difference from the regression line for different algorithms. A low difference indicate a better performance by the metric.

scenes, but scores higher than iCAM for the two other scenes. Totally S-CIELAB scores the highest with 0.68135 for all points, but ΔE^*_{ab} is close with 0.64846 for all points. iCAM scores just below 0.5, and the overall value from SSIM can not be used due to scale problems. All metrics get a lower correlation in the first scene (Figure 72), in the second scene SSIM and iCAM scores significantly higher than for the non-weighted map. For the JP scene all metrics get a higher score, improved with approximately 5 percent. In the last scene SSIM scores more than 0.2 less than in the non-weighted map.

From these results observer regions can improve the correlation between the perceived image difference and the predicted image difference, but only in a number of scenes. In the scenes with an uniform background (Girl and Cartoon) the metrics score lower or approximately the same. For the scenes with a more non-uniform background the results are better, even though S-CIELAB and ΔE^*_{ab} score lower in the Tore scene.



(a) Girl. (b) Tore. (c) Cartoon.

Figure 74: Difference from the regression line - iCAM. Image number 2 is the LM5 image, and we can clearly see that iCAM miscalculates the score.

6.2.3 Eye tracker regions

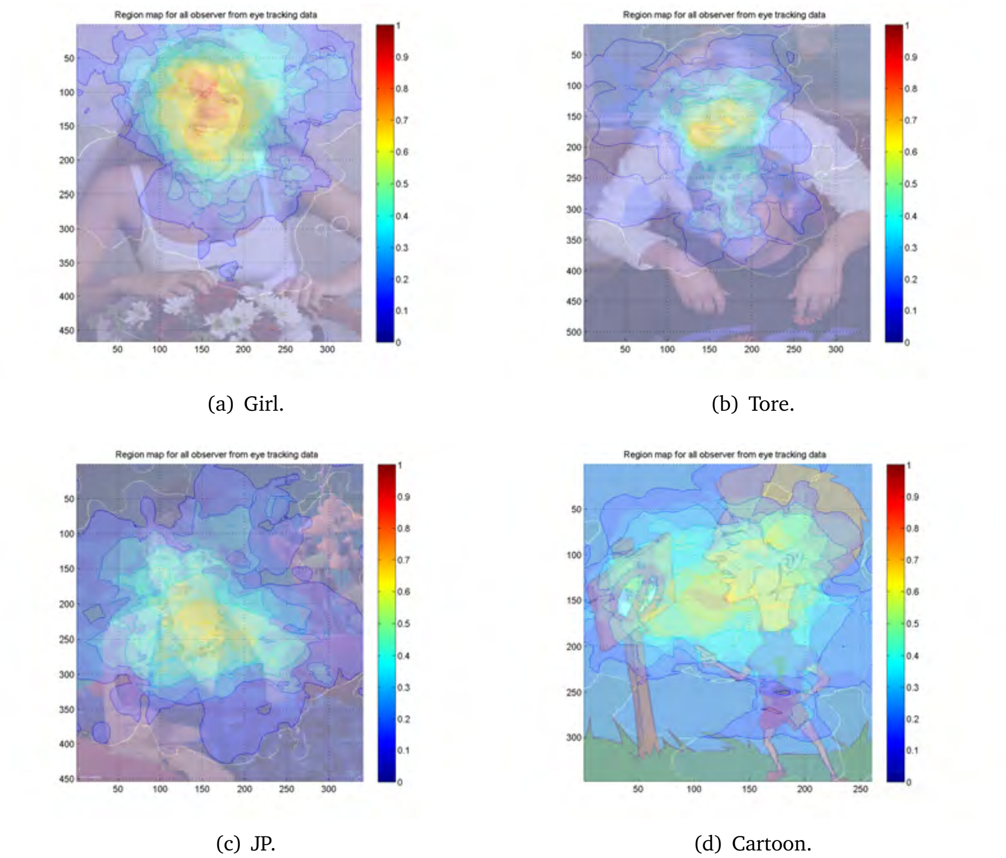
The results from the eye tracker have been used to generate a map (Figure 75) similar to the one from the observer stated regions. This map was based on the frequency map from the eye tracker [34], which was calculated as

$$F(x, y) = \left(\frac{T(x, y)}{N(x, y)} \right)_{NoI} \tag{6.4}$$

where

$$F(x, y) \in [0, 1]$$

The total time an observer used on each pixel was divided by the number of times the observer fixated on that pixel. The result was then normalized by the maximum value in the map [37]. NoI in the formula means normalization. A gaussian filter with width of 35 pixels and height of 35 pixels was then applied to the map to even out differences [33, 34, 37] and to simulate that we look at an area rather than one particular point [82]. For each observer a map was made for each scene, this map had white pixels were the observer had been looking and black for areas with no fixations. The limit set for when an observer had been looking or not were set to 35% of the maximum value. With a threshold of 35% the area match visually across the observer stated regions and the eye tracker map. The results were a map with 1 for a pixel with value greater than 35% of the maximum and 0 for a pixel below this limit, the map was then similar to the one in Figure 66. The overall map was created the same way as for the observer maps by using Equation 6.3 on Page 77.



(a) Girl.

(b) Tore.

(c) JP.

(d) Cartoon.

Figure 75: Binary eye tracker regions for the 4 scenes.

Girl

We can see from Figure 75(a) that most of the areas fixated in the girl scene is located in the face. We do also see that some of the background also have some fixations together with the skin tones on the chest and her left shoulder.

Tore

For this scene we can see from Figure 75(b) that the face of the girl gets the most attention from the observer, we do also see that there are some minor fixations on the guy's face. There are also some attention drawn toward the white sweater and in some background areas.

JP

For this scene the most fixated areas are from the necklace to the lower parts of the face (Figure 75(c)). The most fixated areas are in the same area as the stripe2 (Figure 6(f)). It looks like the necklace worn by the girl could be an important region.

Cartoon

In the Cartoon scene the area with the most fixations are between the sign and the face of the character (Figure 75(d)). The face and some parts of the hair also get a fairly high concentration of fixations. Some parts of the sign also have a high concentration of fixations. This scene has more fixations spread over the entire image than the other scenes.

Overall observervations

We clearly see that the face is an important region, this result has been stated in several research papers [7, 33, 84, 85]. We also see that attention will be drawn toward an eccentric loci [46], an area where a visible change has been made. The scene with the lowest complexity, Cartoon, also has the fixations more spread over the entire scene than the other scenes.

6.2.4 Image difference metrics applied to eye tracker regions

The maps created in the previous section have been applied to image difference metrics.

Scene/Algorithm	S-CIELAB	iCAM	ΔE^*_{ab}	SSIM
1	0.9204	0.6365	0.9210	0.6680
2	0.8969	0.7528	0.9075	0.7605
3	0.1853	0.1563	0.1365	0.1053
4	0.9635	0.7234	0.9559	0.7396

Table 24: r^2 value for the different algorithms, eye tracker regions applied.

S-CIELAB

S-CIELAB gives a fairly linear plot against the z-scores as seen on Figure 76, but it has problems with the Stripe2 images from the JP scene.

In the Girl scene we can see that the two images rated the best and worst from the observers are located at the ends of the S-CIELAB scores as they should. Front LP3 has been rated a little better by S-CIELAB than it should, the value for this should be closer to the values S-CIELAB calculates for LM3 and LP3.

For the Tore scene the two best images are located at the top left corner where they

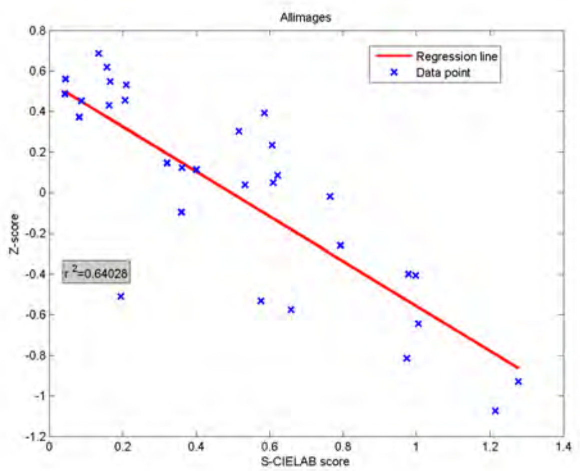


Figure 76: Eye tracker regions: z-score plotted against S-CIELAB values.

should be, and the two images voted to be the worst are given the highest score from S-CIELAB. The only image that S-CIELAB should have given a higher score is the LP3 image.

In the JP scene the two best images are given the lowest score, but the stripe2 images are not given the highest score as they should. Still with only looking at the regions that observers find as important S-CIELAB has problems with predicting the perceived image difference.

S-CIELAB gives good results for the Cartoon scene. The 4 best images with only changed regions, are given the lowest score. The two images, LP3 and LM3 are given approximately the same score and in the same order as the z-score. The two worst images have also been given the highest S-CIELAB score, but not in the order as the z-score.

iCAM

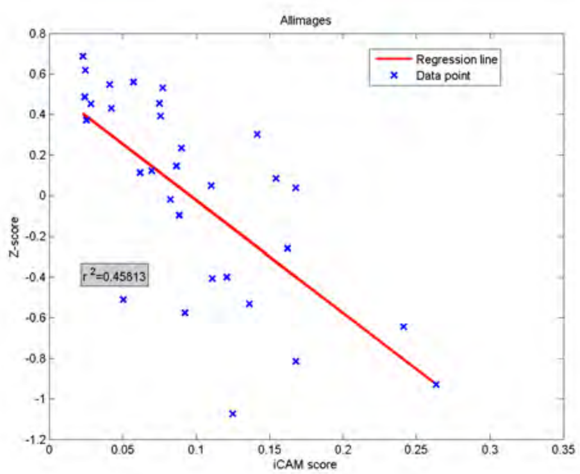


Figure 77: Eye tracker regions: z-score plotted against iCAM values.

iCAM gives us a more scattering plot for the scenes (Figure 77) than for S-CIELAB (Figure 76), there is not a clear linear grouping of the results.

For the Girl scene the two best images are given the lowest score, the fifth best image from the psychophysical experiment is rated as the third best with a large difference from the image that should have been rated as number 3. We do also see that the image rated as the worst by the observers is rated by iCAM as better than 2 other images. It has a z-score lower than -0.8 but iCAM rates it better than one image with z-score of approximately 0.

In the Tore scene the two best images are given the lowest score as they should. The image given third place from the observers are rated as the fifth best image together with image in sixth place. The images with the lowest z-score have only a small iCAM difference from the previous two images.

The JP scene with its regions has been a problem for among other S-CIELAB and we see the same tendency in iCAM. The image with the highest z-score has been rated as fourth in iCAM and the image with the lowest z-score has been rated in third place.

For the Cartoon scene the 4 best images from the psychophysical experiments are rated as the best in the same order by iCAM as the observers. The image stated as the worst by the observers are given an iCAM score more than half of the second to worst image.

ΔE^*_{ab}

For this algorithm we also see a scatter plot as in iCAM, but ΔE^*_{ab} (Figure 78) has a higher correlation coefficient than iCAM (Table 24 and Figure 77). This indicates a better performance by ΔE^*ab than iCAM.

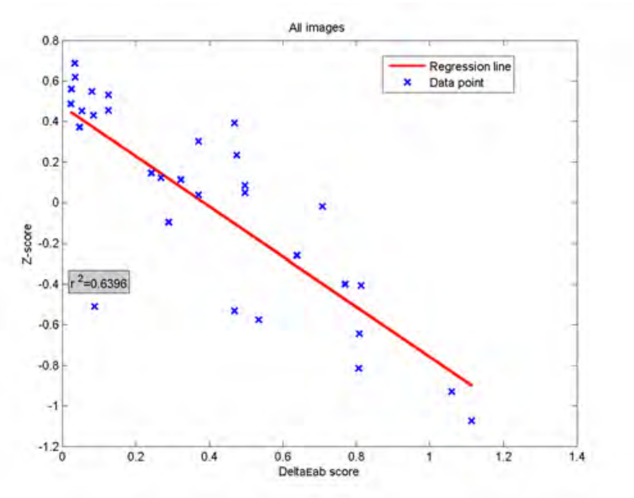


Figure 78: Eye tracker regions: z-score plotted against ΔE^*_{ab} values.

Two of the images in the Girl scene are given a score lower than 0.05 from ΔE^*_{ab}, but from the z-score these values should be higher probably more like 0.5.

For the Tore scene the values from this algorithm are more linear than for the Girl scene. The LP3 image got a too low score, this should have a score closer to LM5 and LP5.

In the JP scene the stripe2 images are given a very low score even though they are stated in lower end of the scale by the observers. They should have been closer to the LM5 and LP5 images in ΔE^*_{ab} values.

For the Cartoon scene we see that ΔE^*_{ab} gives LM3 a better score than LP3, this should have been the other way around.

SSIM

In the Girl scene we can see that SSIM gives the LP5 image a too high score, this should have been closer to the score of LM5. The front LM3 image is calculated as the third worst in SSIM, but the observers stated this as third best (Figure 79).

For the Tore scene the images front LM3 and front LP3 are given a score too low, and the LP3 score should have been lower. The two best images are given approximately the same score, but in the wrong order.

In the JP scene the 4 images rated as the worst from the observers are given SSIM scores all over the scale. The worst image from the psychophysical experiment is given a score close to the score for the two best images. The third best image has also been given a score that should have been higher.

For the Cartoon scene the two best images are given a much lower score than the two other regions even tough there are only small variations in the z-score.

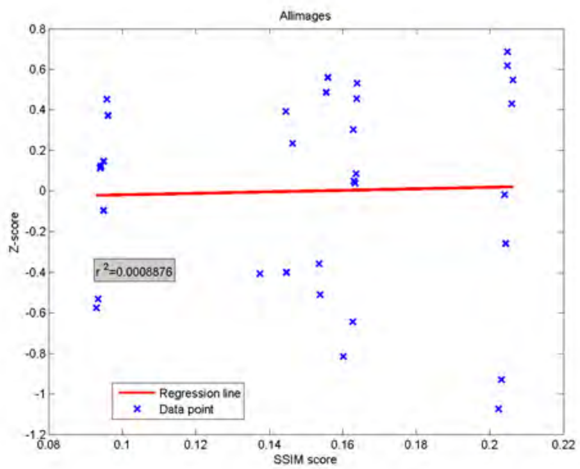


Figure 79: Eye tracker regions: z-score plotted against SSIM values.

Overall algorithm score

We can see from Figure 82 that S-CIELAB scores the lowest, but can not be rated better than iCAM, ΔE^*_{ab}, and it's slightly better than the SSIM algorithm. We do see that iCAM has a lower mean than ΔE^*_{ab}, this was not the case for observer stated regions. For ΔE^*_{ab} it miscalculates with 0.175 to 0.45 for the LM3 images in 3 of 4 scenes, this could be a systematically error. ΔE^*_{ab} could have problems with images that are slightly darker, but when they reach a limit ΔE^*_{ab} can predict perceived difference. iCAM miscalculates the LM5 images, the distance from the regression line to the point is from over 0.8 to 0.4 for 3 out of 4 scenes. iCAM has an improvement in 3 out of 4 scenes, in Tore and JP the correlation is at least 0.15 better than the non-weighted map. SSIM also have improvement in 3 scenes, especially in Tore where the improvement in correlation is more than 0.25. In the third scene all metrics get an increased correlation (Figures 80

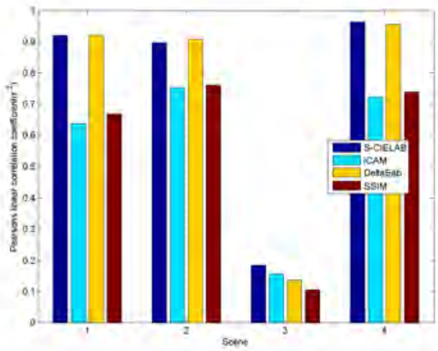


Figure 80: Overview r^2 for observer regions.

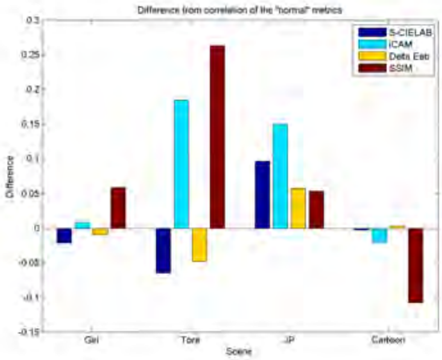


Figure 81: Difference from non-weighted map.

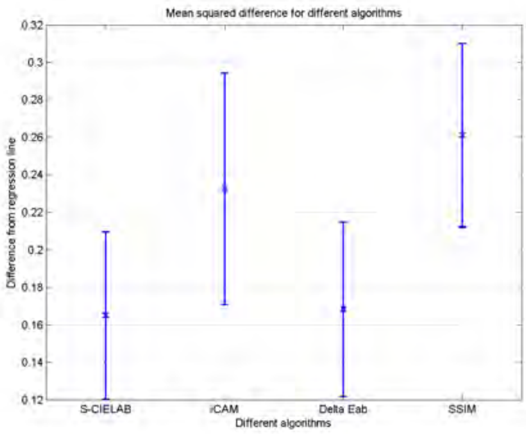


Figure 82: Mean square difference eye tracker regions. A low number indicates a small distance to the regression line.

and 81), while in the last scene SSIM gets a lower correlation and the rest are more or less the same as in the non-weighted computation.

6.2.5 Image difference metrics applied to eye tracker regions - part 2

In this part the fixation maps have been applied to the image difference metrics. The continouos attention maps from the eye tracker have been used (Figure 83). This map has been made the same way as in the previous section but without the threshold to create a binary map. This map is applied to the results from the eye tracker by applying the value from the map for each pixel to the same pixel from the image difference metric (Equation 6.5). By doing it this way the values become smaller, but the correlation coefficient will still be valid as a measure.

$$MetricValue = \frac{\sum_x \sum_y (Mappixel_{x,y} \cdot Metricmap_{x,y})}{N_{numberofpixels}} \quad (6.5)$$

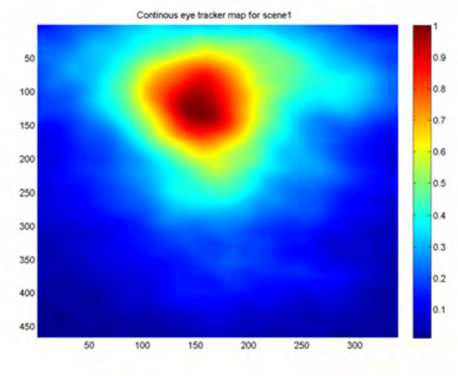


Figure 83: Example continouos eye tracker map.

Scene/Algorithm	S-CIELAB	iCAM	ΔE^*_{ab}	SSIM
1	0.9291	0.6376	0.9270	0.6648
2	0.9230	0.7269	0.9323	0.6856
3	0.1556	0.0791	0.1177	0.0916
4	0.9663	0.7279	0.9552	0.8044

Table 25: r^2 value for the different algorithms, continoues eye tracker regions applied.

S-CIELAB

S-CIELAB gives a plot where the points are located around the regression line. The stripe2 images from the JP scene is the points with the biggest difference from the regression line. This results in a low r^2 score, only 0.1556 for S-CIELAB in this scene (Table 25). The three other scenes score between 0.9230 and 0.9663, which indicates a strong correlation. In the overall score (Figure 84) we can see that the LP5 and LM5 in the Cartoon and Girl scene get different scores, and they are not located as they should be according to the observers.

iCAM

iCAM gets an overall r^2 score of 0.3730, and the points are scattered as seen on Figure 85. The Cartoon scores the best of the scenes in iCAM with 0.7279, but the Tore scene gives

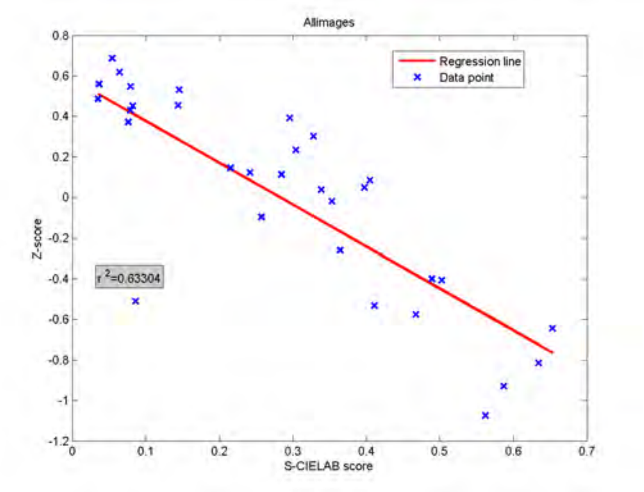


Figure 84: Regression plot for S-CIELAB on continouos eye tracker map.

almost the same score with 0.7269. The biggest problem for iCAM is the LM5 images, iCAM underestimates the score and rates the LM5 as the LM3 and LP3. The same happens in the Girl scene, this results in a r^2 score of 0.6376. For the JP scene none of the points are located near the regression line, which leads to a very low r^2 score.

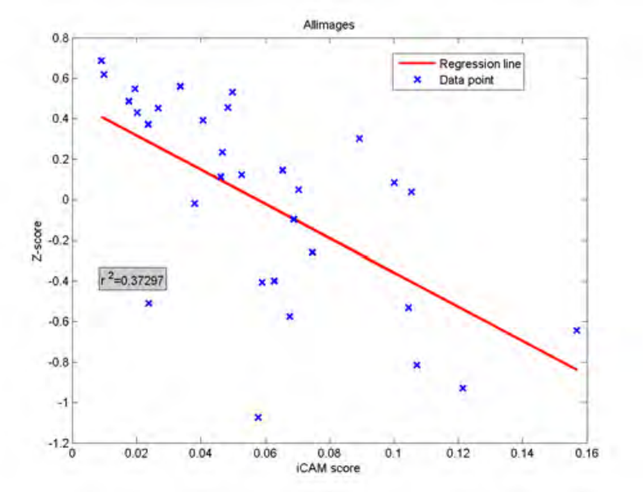


Figure 85: Regression plot for iCAM on continouos eye tracker map.

ΔE^*_{ab}

ΔE^*_{ab} gets an overall score of 0.6496. Some points are located on the regression line (Figure 86), but the stripe2 images are far off the regression line. We do also see that some images are given approximately the same score but the z-score difference is almost 0.5. For the Girl, Tore and Cartoon ΔE^*_{ab} score 0.925, 0.932 and 0.955. This indicate that ΔE^*_{ab} has a good correlation between the z-score and the predicted image difference. The reason for the low overall score is mainly from the JP scene, where we only get a correlation value of 0.118. In the JP scene most of the points are located 0.2 or more off the regression line with the stripe2 images approximately 0.5 and 0.6 from the regression line.

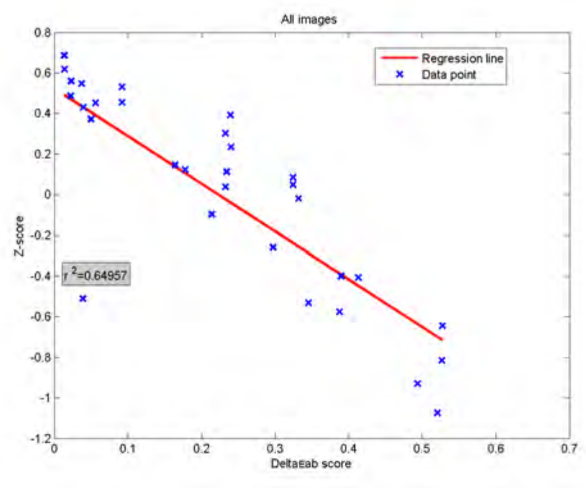


Figure 86: Regression plot for ΔE^*ab on continouos eye tracker map.

SSIM

The overall r^2 value for SSIM is close to 0, this indicate almost no correlation between the points and the regression line. But the scale problem with SSIM makes this value questionable. For the single scenes SSIM shows some correlation. In the Girl scene the r^2 value is 0.66479 and in the Tore scene SSIM scores almost the same. The JP gives a score of almost 0.1, and most of the points are located more than 0.2 from the regression line (Figure 87). Images with more than 0.8 difference in z-score have been rated almost the same. Cartoon is the scene where SSIM scores the highest, with a r^2 of 0.8044. Most of

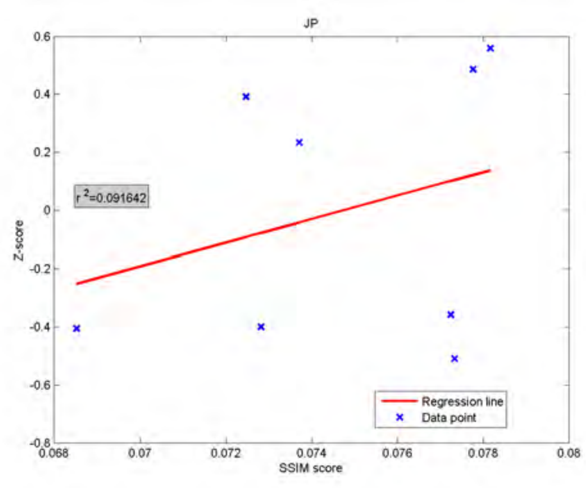


Figure 87: Regression line plot for JP scene with SSIM data points.

the points are located close to the regression line.

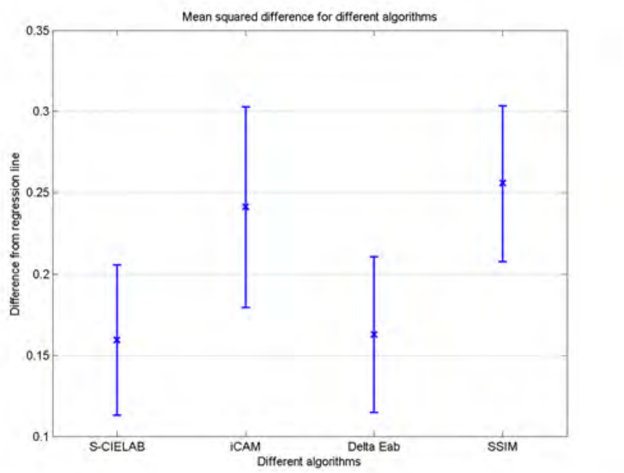


Figure 88: Mean squared difference - eye tracker part 2.

Overall score for metrics

We can see from Figure 89 that S-CIELAB and ΔE^*_{ab} score approximately the same in each scene, while iCAM and SSIM are more similar in their r^2 score. For the JP scene all metrics score low, but S-CIELAB has the highest value while iCAM has the lowest. SSIM gets a higher correlation in three first scenes with a weighted eye tracker map, while in the last scene it scores a little lower than the non-weighted (Figure 90). Both iCAM and SSIM improve their correlation for the Tore scene. All metrics improve in the third scene, probably because the center of the eye tracker map is located in the same area as the horizontal stripe.

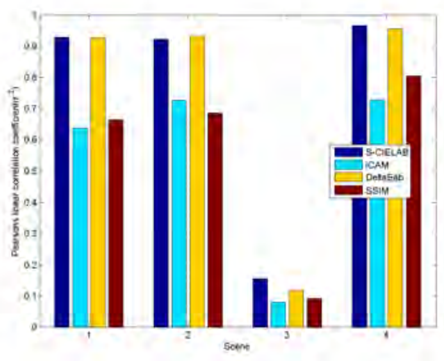


Figure 89: Overview r^2 for eye tracker part 2.

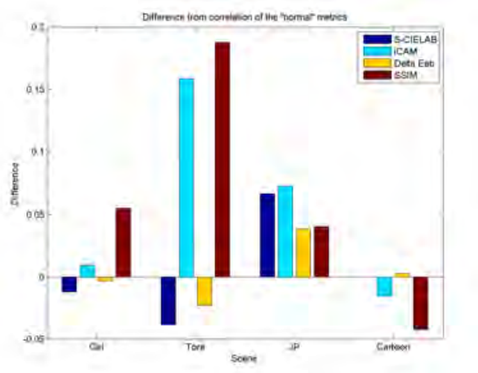


Figure 90: Difference from non-weighted map.

If we look at the mean squared difference from the regression line on Figure 88, S-CIELAB and ΔE^*_{ab} have almost the same mean. iCAM and SSIM have a higher mean value, but iCAM cannot be differeniated from the rest. S-CIELAB is better than SSIM because the confidence intervals do not overlap.

From the results S-CIELAB and ΔE^*_{ab} seem to be the best performing algorithms when we take into account the correlation coefficient.

6.2.6 Image difference metrics applied to a combined observer and eye tracker map

We have combined the maps from Section 6.2.2 and 6.2.4 to create a mean observer and eye tracker map. There are differences between these maps, and a combination of these will cover both what the observer is looking at and what he or her states as important. The map used was created by taking the mean observer map for each scene and adding the mean eye tracker map. The resulting map was divided by 2 for normalization. The value from the metric was then computed the same way as in Equation 6.5.

Scene/Algorithm	S-CIELAB	iCAM	ΔE^*ab	SSIM
1	0.9366	0.5988	0.9142	0.6196
2	0.9300	0.7309	0.9312	0.8905
3	0.1623	0.0995	0.1238	0.1061
4	0.9624	0.7426	0.9506	0.8489

Table 26: r^2 value for the different algorithms, combined regions applied.

S-CIELAB

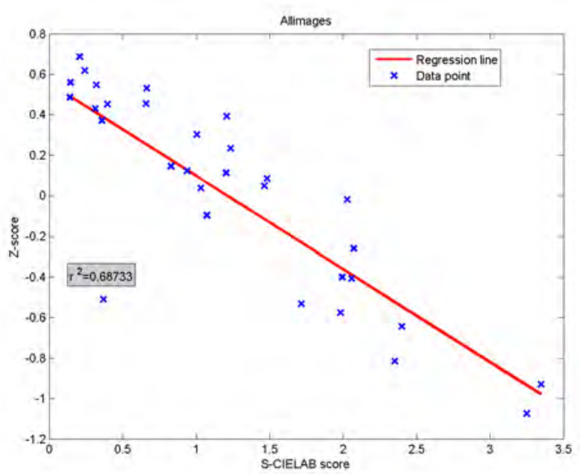


Figure 91: Regression plot for S-CIELAB on combined eyetracker and observer map.

S-CIELAB gives an overall r^2 score of 0.68733 (Table 26), most of the data points are located around the regression line except the JP stripe2 images (Figure 91). For the single scenes the Girl, Tore and Cartoon score similar with values from 0.9300 to 0.9624. This indicates a good correlation between the computed values and the observer values. For the JP scene we get a score a little above 0.16, and here the stripe2 images have the largest difference from the regression line.

iCAM

iCAM scores 0.48945 for the overall r^2 value. Most of the values with a z-score higher than 0 are located close to the regression line, but for the values with a z-score lower than 0 this is not the case (Figure 92). For the Girl scene the r^2 value is 0.59881, and the image farthest off the regression line is the LM5. This image is also the farthest off the regression line in the Tore scene, but here the r^2 value is higher with 0.7309. The

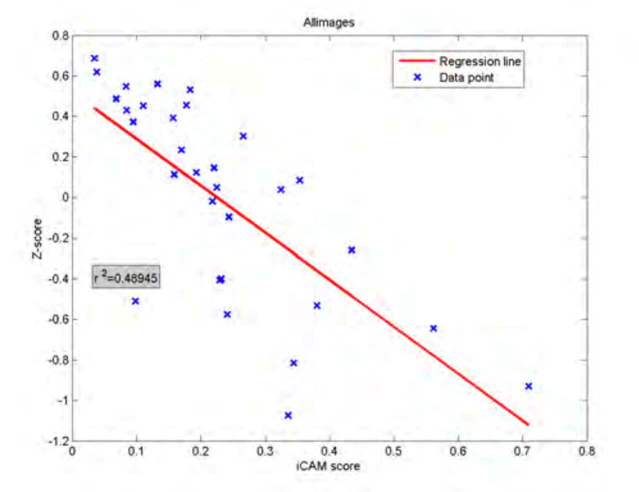


Figure 92: Regression plot for iCAM on combined eyetracker and observer map.

Cartoon scene gets a score a little better than Tore, but we also see the problem with the LM5 image here. In the JP scene the points are located far off the regression line, and the r^2 value also indicates a very weak correlation.

ΔE^*_{ab}

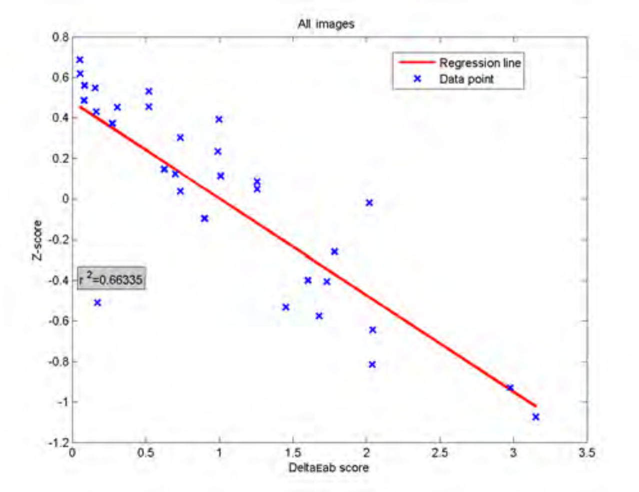


Figure 93: Regression plot for ΔE^*_{ab} on combined eyetracker and observer map.

The overall score for the ΔE^*_{ab} is 0.66335, most of the points are located around the regression line as seen on Figure 93. There is also a problem here when it comes to the JP stripe2 images. The Girl, Tore and Cartoon scenes all have a r^2 value greater than 0.9. This indicate a high correlation between the predicted image difference and the observer scores. For the JP scene most of the values are located off the regression line, the stripe2 images have the largest difference and the r^2 is 0.12383.

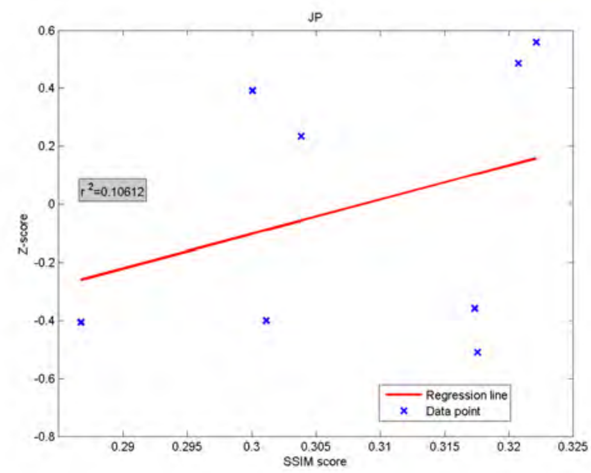


Figure 94: Regression plot for SSIM on combined eyetracker and observer map on JP scene.

SSIM

The SSIM gets an overall r^2 close to 0, but as explained earlier this is not a valid measurement. The r^2 makes more sense when used on a single scene. For the Girl scene we get a correlation coefficient of 0.6196. SSIM has some problemes here, the LP5 has been rated almost as good as the LM3, but the z-score distance between these are almost 0.8. For the Tore scene we get a score of 0.8905, this indicate a high correlation between the predicted image difference and the observer scores. In the JP scene SSIM only scores 0.1061. 4 images have been rated in the lower end of the scale by the observers, but in SSIM these are spread from third best to worst (Figure 94). Due to this SSIM shows a low correlation for this scene. The Cartoon scene scores 0.8489, here most of the points are located close to the regression line and therefore the high correlation.

Overall score for metrics

If we take a look at the overall scores for the different metrics in Figure 95, S-CIELAB and ΔE^*_{ab} have the lowest mean but due to the confidence intervals these cannot be differeniated from the iCAM or the SSIM. From Figure 96 we can see that S-CIELAB and ΔE^*_{ab} have the highest score for all scenes, this is in connection with the mean squared difference from the regression line where these have the lowest mean. The SSIM score is close to the S-CIELAB and ΔE^*_{ab} especially in the Cartoon scene, but also in Girl. For the JP scene all metrics score low, but S-CIELAB is also the best here. iCAM scores the worst in all scenes, but in JP and Tore there is not much separating this metric from SSIM.

SSIM is the metric with the highest improvement over the non-weighted metric (Figure 97), SSIM scores better in all scenes and improves with 0.4 for the Tore scene. In the same scene iCAM also gets a higher score. All metrics improve in the third scene.

6.2.7 Image difference metrics applied to freeview

The frequency map from the freeview has been applied to the metrics. In a freeview task observers will have another gaze area than if they are given a specific task [59, 64]. By applyping the frequency map from the freeview we can check if we improve the correlation between the perceived image difference and the predicted image difference.

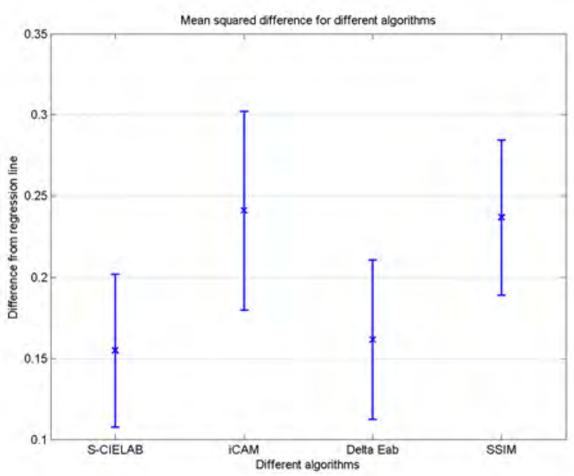


Figure 95: Mean squared difference from regression line for combined eye tracker and observer map.

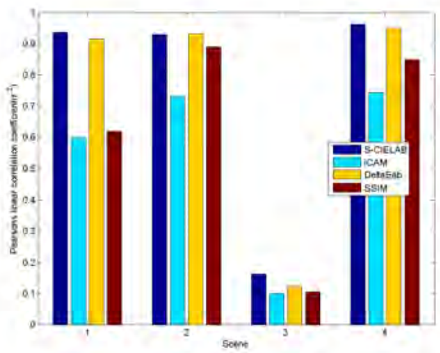


Figure 96: Overview r^2 for combined eye tracker and observer map.

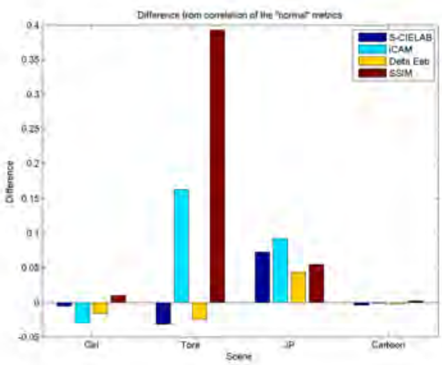


Figure 97: Difference from non-weighted map.

Scene/Algorithm	S-CIELAB	iCAM	DeltaEab	SSIM
1	0.8288	0.6017	0.8315	0.8054
2	0.8611	0.6823	0.8777	0.5306
3	0.0631	0.0014	0.0662	0.0339
4	0.9670	0.7290	0.9580	0.7870

Table 27: r^2 value for the different algorithms, freeview regions applied.

S-CIELAB

The S-CIELAB gives an overall correlation of 0.553 for the 4 scenes (Table 27 and Figure 98). In Girl, Tore and Cartoon we get a correlation above 0.8. In the JP scene we only get a correlation of 0.063, mainly because of the images with the horizontal stripe.

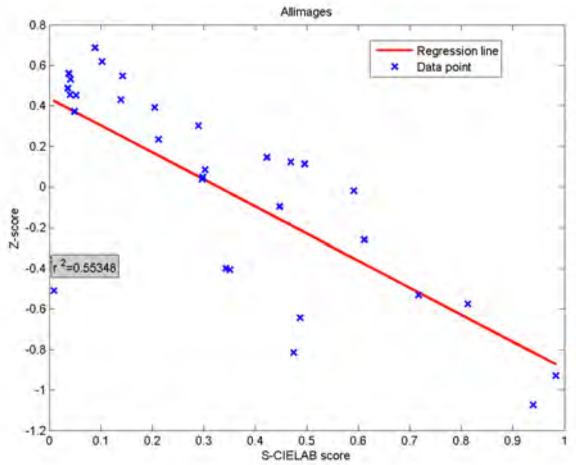


Figure 98: Freeview - regression plot S-CIELAB.

iCAM

With a score just below 0.4 iCAM do not predict the perceived image difference very well (Figure 99). The images with LM5 are still miscalculated and draw the correlation down.

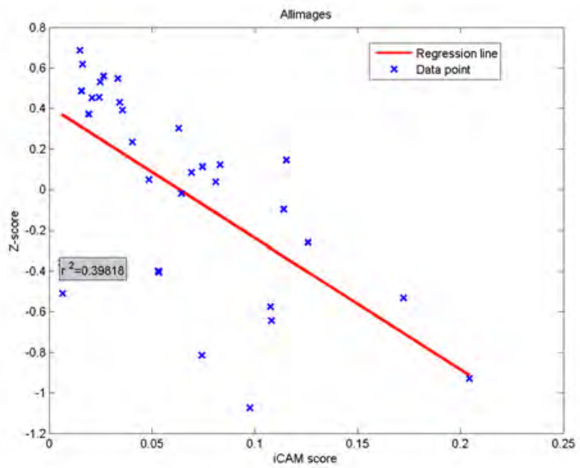


Figure 99: Freeview - regression plot iCAM.

ΔE^*_{ab}

This metric gets an overall correlation of 0.57195 (Figure 100). We still see that ΔE^*_{ab} has problems with the horizontal stripes in the JP scene. In this scene ΔE^*_{ab} only scores 0.066, while in the other scenes we get a score between 0.8315 and 0.9580.

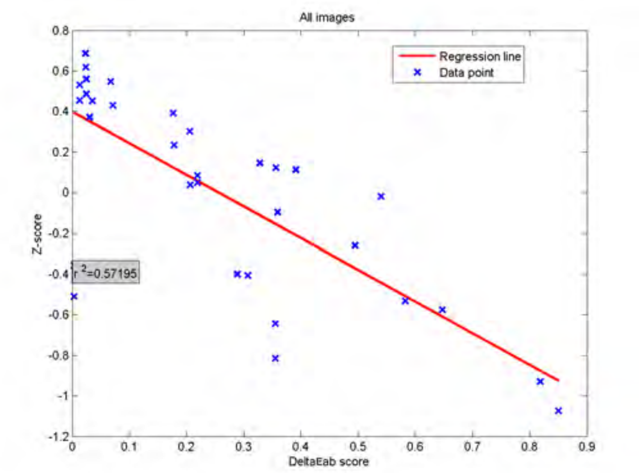


Figure 100: Freeview - regression plot ΔE^*_{ab}.

SSIM

SSIM gives an overall score very close to 0 (Figure 101), while the different scenes indicate a better correlation. In the Cartoon scene and Girl scene SSIM scores a correlation of 0.787 and 0.805. In the JP scene SSIM does not predict the perceived image difference well, and the correlation is only 0.034. In the Tore scene we get a moderate correlation with 0.53.

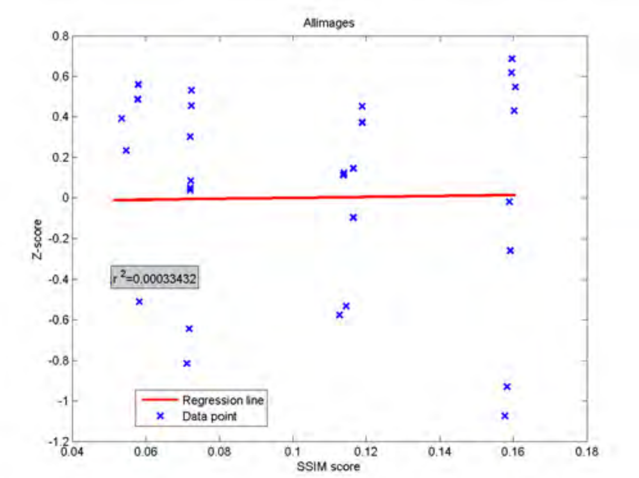


Figure 101: Freeview - regression plot SSIM.

Overall score for metrics

From Figure 102 there is no way of telling which metric is the best, but S-CIELAB and ΔE^*_{ab} have the lowest mean. For the correlation S-CIELAB and ΔE^*_{ab} have the highest correlation over the 4 scenes (Figure 103), but in the first scene SSIM has been improved with almost 0.2 and compete with S-CIELAB and ΔE^*_{ab} here. From Figure 104 we can see that most of the metrics in the scenes scores worse or the same as the non-weighted. From this we cannot say that a freeview weighting of the metrics will improve the correlation between the perceived and predicted image difference.

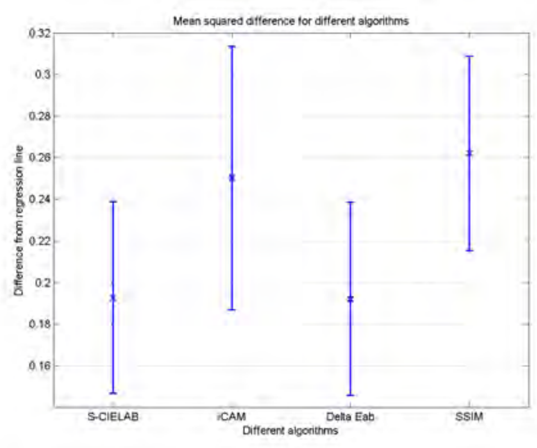


Figure 102: Mean squared difference from regression line for the freeview task.

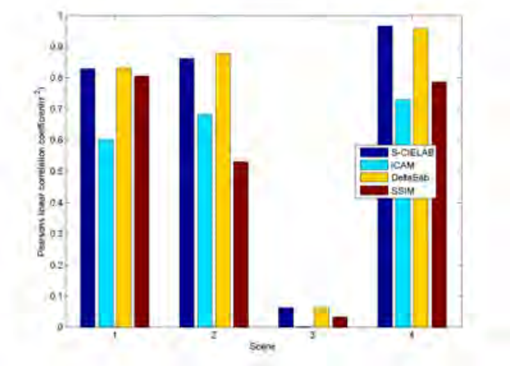


Figure 103: Overview r^2 for freeview.

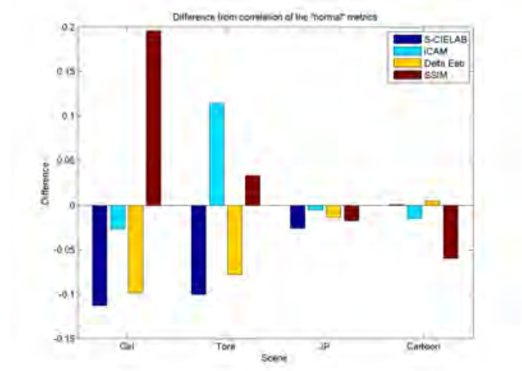


Figure 104: Difference from non-weighted map - Freeview.

6.2.8 Image difference metrics applied to gaze marking task

The frequency maps from the gaze marking, where the observers were asked to look at regions important for their choice, are applied to the image difference metrics.

S-CIELAB

S-CIELAB gives a correlation coefficient of 0.593 (Table 28 and Figure 105), the images with a low S-CIELAB score except JP stripe are located in a cluster at the top left corner near the regression line. The other images are more spread and have a greater difference from the regression line, but overall there is a correlation between the predicted image difference and the perceived image difference.

iCAM

For the iCAM metric the overall correlation is 0.378. From Figure 106 we can see that the points are scattered, the image rated the worst by the observers is located in the better half by iCAM.

Scene/Algorithm	S-CIELAB	iCAM	DeltaEab	SSIM
1	0.8288	0.6017	0.8315	0.8054
2	0.8611	0.6823	0.8777	0.5306
3	0.0631	0.0014	0.0662	0.0339
4	0.9670	0.7290	0.9580	0.7870

Table 28: r^2 value for the different algorithms, gaze marked regions applied.

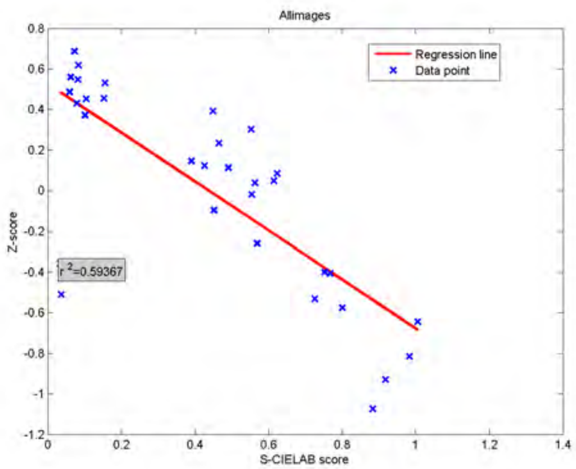


Figure 105: Gaze marking - regression plot S-CIELAB.

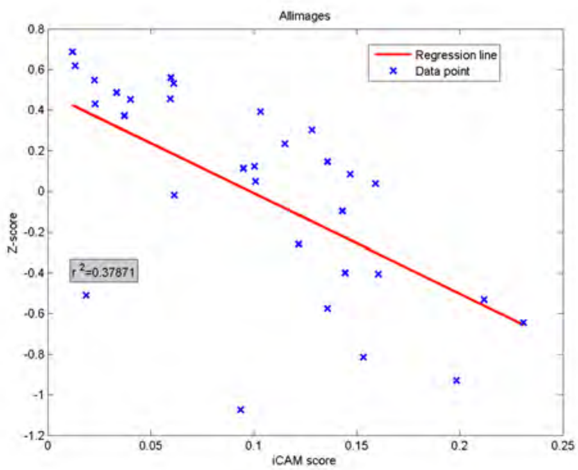


Figure 106: Gaze marking - regression plot iCAM.

ΔE^*_{ab}

The overall correlation for this metric s 0.6125. The JP stripe images have the greatest distance from the regression line. We do also see that we have a cluster of points in the images rated the best by the metric, but the other images are more scattered than this cluster (Figure 107).

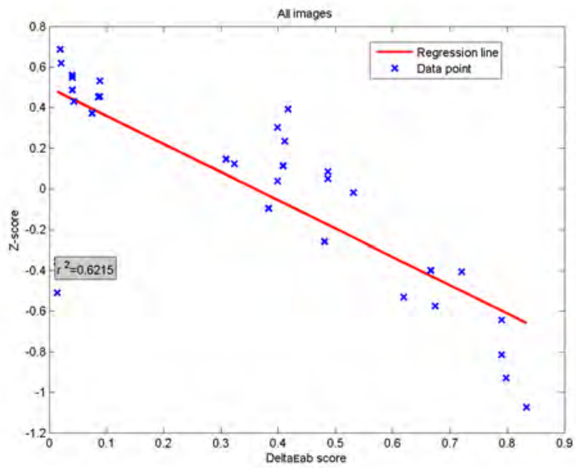


Figure 107: Gaze marking - regression plot ΔE^*_{ab}.

SSIM

Overall score here is close to 0 (Figure 108), but for the different scenes we have a correlation. For the Cartoon the correlation is 0.814, which indicate a strong correlation. For the Girl and Tore scene we have a correlation of 0.677 and 0.538, while in the JP scene the correlation is only 0.0419.

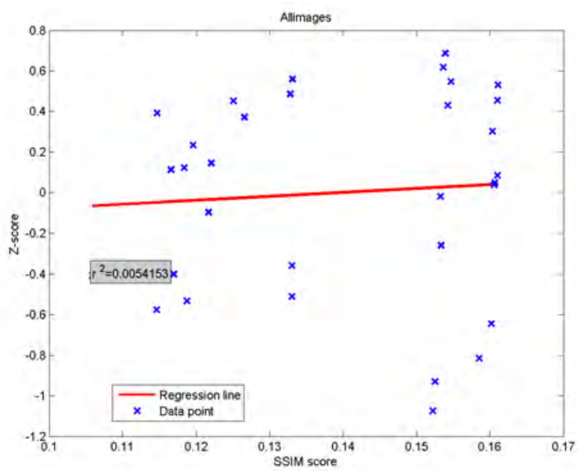


Figure 108: Gaze marking - regression plot SSIM.

Overall score for metrics

From the results S-CIELAB and ΔE^*_{ab} have the lowest mean squared difference from the regression line (Figure 109), but cannot be differeniated from the SSIM and iCAM. In the correlation plot (Figure 110) S-CIELAB and ΔE^*_{ab} have the highest score through all of the 4 scenes, but these two metrics have not been improved compared to the non-weighted version of the metrics. SSIM has been improved in the first and second scene, but score slightly lower in the two last scenes (Figure 111). iCAM improves by 0.15 in the second scene, but in the three other scenes the difference is minor.

According to this gaze marking weighting of the metrics does not improve the overall performance of the metrics, but in some scenes an improvement is found.

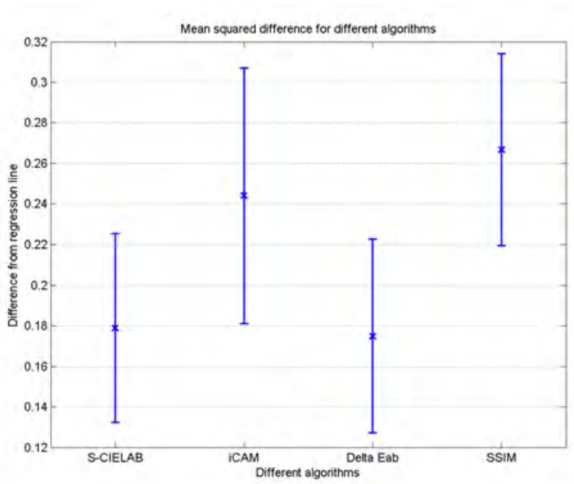


Figure 109: Mean squared difference from regression line for gaze marking.

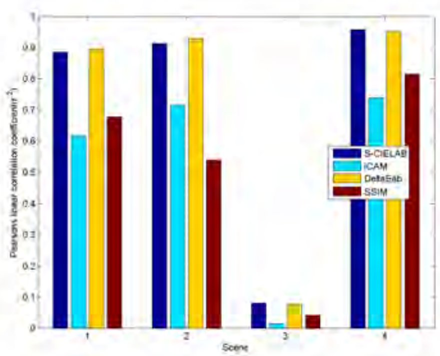


Figure 110: Overview r^2 for gaze marking.

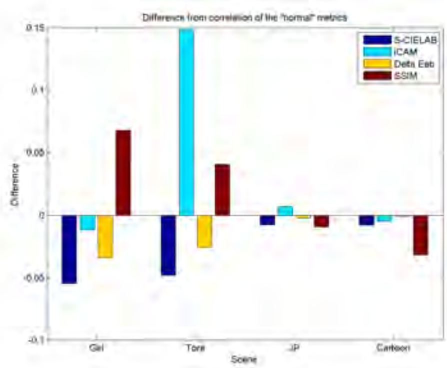


Figure 111: Difference from non-weighted map - gaze marking.

6.2.9 Gaze maps from similar images applied to image difference metrics

For the different scenes image pairs with similar modifications have been extracted together with their frequency maps from the eye tracker. The frequency maps are computed in the same way as in Section 6.2.5. For the JP scene the LM3 and LP3, LM5 and LP5, stripe LM3 and stripe LP3 and stripe2 LM3 and stripe2 LP3 regions have been rated similar, and for all valid observers their gaze positions have been extracted for each image. This comes to a totalt of 8 image pairs for each of the observers. The JP scene have the lowest overall correlation among the scenes for all metrics, therefore this scene has been used to check if maps from similar images can provide an improvement for the metrics.

For S-CIELAB, iCAM and ΔE^*_{ab} we only get minor improvements in the correlation (Figure 112), but for the SSIM we get a correlation of 0.17. This is an improvement from the 0.05 in the normal computation way. The horizontal stripe images have been rated lower than the vertical stripes, this "correct" rating is the main reason for the improvement. Even though we get an improvement of 0.11 the correlation for SSIM in this scene is still very low, indicating that nor SSIM or the other metrics can predict perceived image difference very well for this scene even with a weighting of this kind.

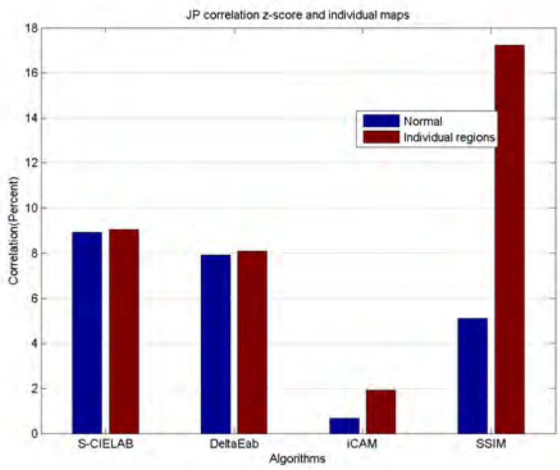


Figure 112: Correlation for individual gaze maps in JP scene.

6.2.10 Evaluating the different maps and their effect on image difference

Each algorithm has been applied to the different maps (Observer, eye tracker binary, a combination of these and the eye tracker continouos, freeview and gaze marking). We will now evaluate the differences between these maps. Table 29 shows the mean squared distance from the regression line for the different algorithms over the different maps.

Metric/Map	Normal	Observer	Eye tracker(B)	Eye tracker/Observer	Eye tracker(C)	Freeview	Gaze marking
S-CIELAB	0.156	0.165	0.165	0.155	0.159	0.193	0.179
iCAM	0.251	0.249	0.233	0.241	0.241	0.250	0.244
Delta Eab	0.162	0.179	0.168	0.162	0.163	0.192	0.175
SSIM	0.269	0.273	0.261	0.237	0.256	0.262	0.267

Table 29: Mean squared distance from regression line.

We can see that S-CIELAB has the lowest distance from the regression line in all maps except freeview and gaze marking, and there are only minor changes in the mean squared difference. The same goes for ΔE^*_{ab}. iCAM has a distance of 0.251 in the normal computation of the perceived image difference, when applying the binary eye tracking map the value decreases to 0.233. In SSIM the normal way of computing the image difference gets 0.269, while applying a combination of the binary eye tracking map and the observer maps the value decreases to 0.237.

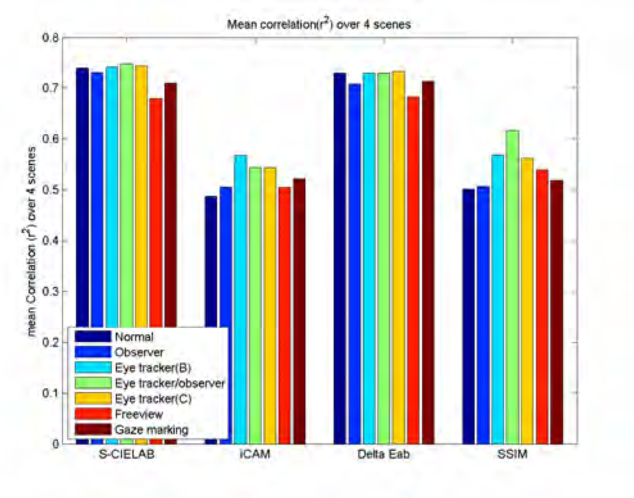


Figure 113: Mean correlation over 4 scenes.

If we take a look at the other measures of evaluating the results, the r^2 correlation as seen in Figure 113 and Table 30, S-CIELAB and ΔE^*_{ab} are more or less stable over the 5 first maps, with some larger differences in freeview and gaze marking. In iCAM the eye tracker binary maps gets the highest correlation. It is worth noticing here that all the different maps scored a higher r^2 than the normal map. The same goes for SSIM, but here the combination map between eye tracking and observer scored the best.

Metric/Map	Normal	Observer	Eye tracker(B)	Eye tracker/Observer	Eye tracker(C)	Freeview	Gaze marking
S-CIELAB	0.740	0.731	0.742	0.748	0.743	0.680	0.710
iCAM	0.487	0.505	0.567	0.543	0.543	0.504	0.521
Delta Eab	0.729	0.708	0.730	0.730	0.733	0.683	0.714
SSIM	0.501	0.506	0.568	0.616	0.562	0.539	0.518

Table 30: Mean correlation over 4 scenes.

S-CIELAB

In S-CIELAB we had an overall good correlation in scenes 1, 2 and 4 (Girl, Tore and Cartoon) with only minor changes between the maps (Figure 114). In scene 3, JP, the normal map scored the worst and the binary eye tracking map scored the best. The map from the observers have a high weighting on the upper part of the body (Figure 75(c)), this is where the horziontal stripe is located and this could explan why this map improves the correlation. We do also see that all maps score better than the normal. This is probably due to the regions that have been altered in this scene. Over the 4 scenes the combination of eye tracker binary map and the observer map gives the best correlation, this could indicate an improvement for S-CIELAB when you take into account both gaze information and subjective information from the observers.

iCAM

For iCAM we have more variations in the data as seen on Figure 115. For the first scene, Girl, the observer maps have the highest correlation but there are only small changes to

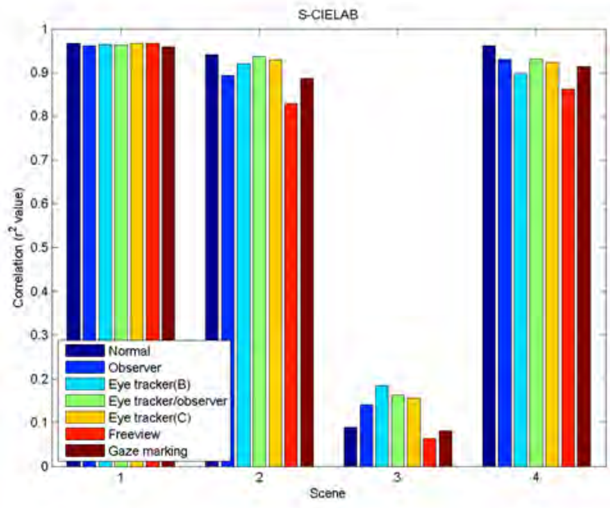


Figure 114: Correlation - S-CIELAB.

the other maps. In the second scene the observer map is rated as the worst while the binary eye tracking map is rated as the best, but only minor difference to the other maps. In the third scene, JP, the binary eye tracking map gets the highest r^2 value while observer map and the normal have the worst score. In the last scene the binary eye tracking map is rated as the best with the other maps close except the normal. The normal scores significantly lower than the rest. In Figure 113 we do also see that the eye tracker binary

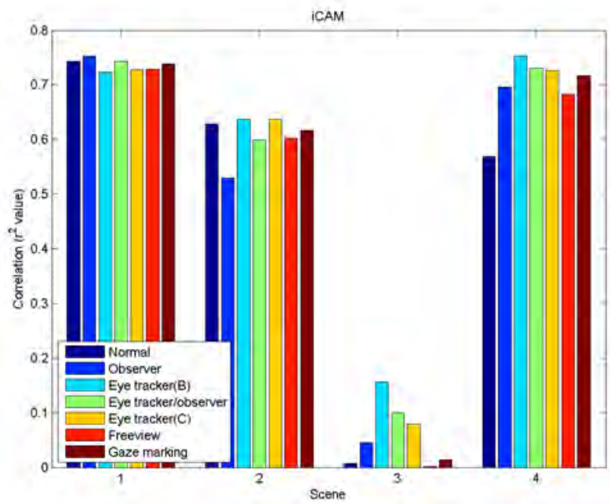


Figure 115: Correlation - iCAM.

(Eye tracker (B)) has the highest mean over the 4 scenes.

ΔE^*_{ab}

For the ΔE^*_{ab} metric we get more or less a similar results for the first scene (Figure 116). For the second scene only the observer map scores lower than the rest. This could be because the observers have mainly stated the faces as important without marking a lot of the background. Some images here have only been altered in the background, these are then given a higher score by the metric than they should have. In the JP scene we have a low correlation for all maps, but the binary eye tracking map scores the best and the normal map has the lowest correlation. The choice of regions and modifications to these regions resulted in problems for all metrics, including ΔE^*_{ab}. But we do see that regions improve the performance of ΔE^*_{ab} in this scene. In the last scene, Cartoon, normal computation has the highest correlation, while the rest are more or less similar with only minor differences.

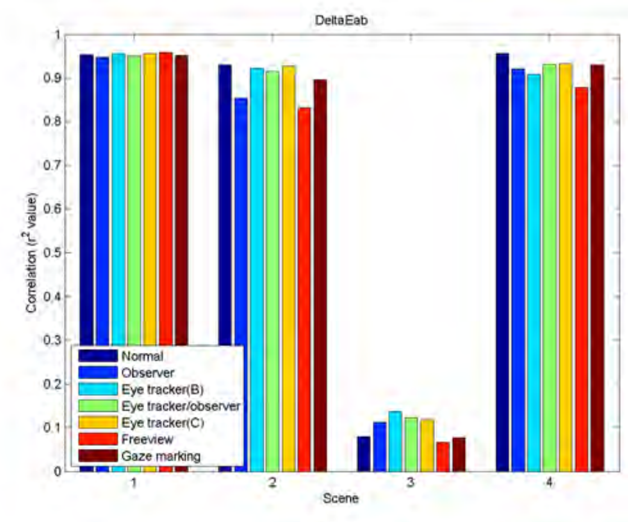


Figure 116: Correlation - ΔE^*_{ab}.

SSIM

SSIM shows more variation in the different scenes than the other metrics (Figure 117). In the first scene the normal and combined eye tracker binary and observer map are rated as the best, while observer regions are rated as the worst. In the second scene, Tore, the freeview map is clearly the best, while observer regions is rated as the worst. In the JP scene the freeview map is rated as the worst, while the eye tracker and observer map score almost the same. The reason why the normal also gets a low correlation here is probably due to the small very visible regions. In the last scene the combined eye tracker and observer map gets the highest rating while the normal map has a score almost half of the best.

Overall observations

For the JP scene where we have a region with a small change, but still very visible, different maps can improve the performance of image difference metrics. In images with visible, but not necessary a large difference, eye tracking and/or observer maps can improve the performance of todays image difference metrics. Overall we see an improvement in the metrics, especially in metrics with a correlation in the normal way of computation. One wheighting map cannot be rated as better than other, this is both metric and scene dependent.

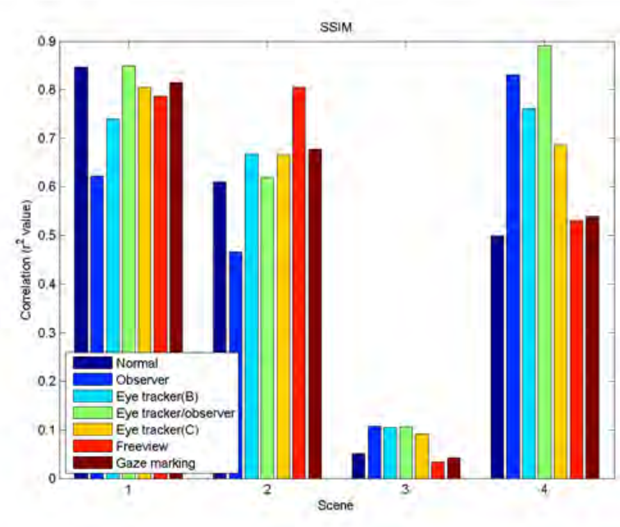


Figure 117: Correlation - SSIM.

7 Conclusion

Overall the objectives of this project have been met. The first goal was to investigate the importance of region-of-interest on image difference metrics. The second goal was to investigate how observers looked at images during different tasks, what kind of differences there are between not only tasks, but also different groups.

7.1 How do we look at images

The analysis of how we look at images in Chapter 5 has provided useful information. The results indicate a difference between experts and non-experts both in how they look at images and what they mark as important. Non-experts look at and mark larger areas, the entire appearance is more important. For expert observers details and preciseness is more important.

Results also indicate variations between the different tasks given to the observers. In a freeview task faces are the most important regions while in a pair comparision task the gaze position is shifted to other parts where the observers see a difference or have a reference point. Gaze marking important regions do not correlate well with observers manually marked regions, indicating that gaze marking cannot substitute marking on paper with a pen.

Observers tend to rate images visually similar different than images visually different. In similar images the observers tend to choose one of the sides when picking the image most similar to the original. Observers also use the original image actively during the whole process, indicating a high value while rating and that reproduction to reproduction evaluation is less important then reproduction to original evaluation.

Then to answer the research question stated in the introduction "How do observers look at images given different tasks?", observers will change the way they look at images together with the task given to them. Faces are clearly an important area, together with areas with semantic value. Differences are also found in what experts and non-experts find important in rating image difference.

7.2 Image difference metrics

In Chapter 6 different image difference metrics were evaluated. The metrics based on ΔE^*_{ab} showed the best results in the normal computation, with indications of the hue angle metric to perform the best.

Different area based image difference were computed, using information from the different tasks given to the observers. The results indicate that area based image difference can improve image difference metrics to a certain degree. Metrics with a high correlation to the perceived image difference have little or non improvements, while in metrics with a low correlation area based image difference will increase the performance. Area based image difference directly from the psychophysical experiment (eye tracker data or observer marked regions) seem to give the best results, but this is both scene and metric dependent.

The research question stated in the beginning of thesis was ”Can region-of-interest improve overall image difference metrics in complex images?”. Image difference metrics can be improved with region-of-interest, but the region-of-interest that should be used is both metric and scene dependent.

8 Further research

The most obvious direction for future work would be to expand this research to include gamut mapped images. In gamut mapped images highly visible regions can occur and a gaze map weighting can clarify the use of gaze maps in image difference metrics in gamut mapping evaluation. By performing this on gamut mapped images may give more information about what kind of areas and differences we look at can be discovered and quantified.

Another way of extending this research would be by using different scenes, not only portrait images. There is also several other metrics that could be implemented and tested. We have only used one kind of heatmap in this research, the frequency map proposed by Bai et al. [34]. Other ways of computing these maps and applying them to image difference metrics will be an interesting research project.

The research can also be expanded to use region-of-interest algorithms instead of eye tracking. Eye tracking is both time consuming and resource demanding and cannot be carried out for every image. With the improvement found eye tracking can be replaced with algorithms simulating human region-of-interest, this would be a natural way for directing further research. Region-of-interest algorithms can also be computed into the framework of an image difference metrics, which can be considered in furher work.

In this research project only digital images on screen has been considered, but this can also be extended to include other medias as print.

There were also found changes in time used for different groups, a natural way of expanding this would be to see how time elapses in different reproductions and over several groups. We have also shown the difference between experts and non-experts in marking and looking at images, it would be natural to expand this to find out more information about the differences between these groups.

Bibliography

[1] John M. Henderson, Carrick C. Williams, Monica S. Castelhano, and Richard J. Falk. Eye movements and picture processing during recognition. *Perception & Psychophysics*, 65:725–734, 2003.

[2] Alejandro Jaimes, Jeff B. Pelz, Tim Grabowski, Jason S. Babcock, and Shih-Fu Chang. Using human observer eye movements in automatic image classifiers. In *Proceedings of SPIE: Human Vision and Electronic Imaging VI*, volume 4299, pages 373–384, 2001.

[3] Geoffrey Underwood and Tom Foulsham. Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *The Quarterly Journal of Experimental Psychology*, 59:1931 – 1949, November 2006.

[4] Umesh Rajashekar, Lawrence K. Cormack, and Alan C. Bovik. Image features that draw fixations. *IEEE International Conference on Image Processing*, 3:313–316, 2003.

[5] Jason S. Babcock, Marianne Lipps, and Jeff B. Pelz. How people look at pictures before, during, and after scene capture: Buswell revisited. *Proc. SPIE Human Vision and Electronic Imaging VII*, 4662:34–47, 2002.

[6] Umesh Rajashekar, Lawrence K. Cormack, and Alan C. Bovik. Point of gaze analysis reveals visual search strategies. *Human Vision and Electronic Imaging*, 5292:296–306, 2004.

[7] Chizuko Endo, Takuya Asada, Hideaki Haneishi, and Yoichi Miyake. Analysis of the eye movements and its applications to image evaluation. In *IS&T and SID's 2nd Color Imaging Conference: Color Science, Systems and Applications*, pages 153–155, 1994.

[8] Norman H. Mackworth and Anthony J. Morandi. The gaze selects informative details with pictures. *Perception & psychophyscics*, 2:547–552, 1967.

[9] Richard E. Parker. Picture processing during recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 4(2):284–293, 1978.

[10] David E. Irwin. *Eye movements and visual cognition: scene perception and reading*, chapter Visual memory within and across fixations, pages 146–165. Springer Series in Neuropsychology. Springer, New York, 1992.

[11] Eriko Bando, Jon Y. Hardeberg, David Connah, and Ivar Farup. Prediction visible image degradation by colour image difference fomulae. In *The 5th International Conference on Imaging Science and Hardcopy*, pages 121–124, 2004. ISSN 0806-3176.

[12] Nicolas Bonnier, Francis Schmitt, Hans Brettel, and Stephane Berche. Evaluation of spatial gamut mapping algorithms. In *Fourteenth Color Imaging Conference*, volume 14, pages 56–61, November 2006. ISBN / ISSN: 0-89208-292-5.

[13] Gaurav Sharma et.al. *Digital Color Imaging Handbook*. CRC Press, 2002. ISBN: 084930900X.

[14] Guowei Hong and M. Ronnier Luo. Perceptually based colour difference for complex images. In Allan Rodrigues Robert Chung, editor, *Proceedings of SPIE: 9th Congress of the International Colour Association*, volume 4421, pages 618–621, 2002.

[15] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[16] Xuemei Zhang and Brian A. Wandell. A spatial extension of CIELAB for digital color image reproduction. *SID Journal*, 1997. 21/05/05: http://white.stanford.edu/~brian/scielab/references.html.

[17] Maria Vanrell, Ramon Baldrich, Anna Salvatella, Robert Benavente, and Francesc Tous. Induction operators for a computational colour-texture representation. *Computer Vision and Image Understanding*, 94(1-3):92–114, April-June 2004.

[18] Xuemei Zhang. Introduction to S-CIELAB. 21/05/07: http://white.stanford.edu/ brian/scielab/scielab.html, 1998.

[19] Jiangtao Kuang, Garrett M. Johnson, and Mark D. Fairchild. iCAM for high-dynamic-range image rendering. In *APGV '06: Proceedings of the 3rd symposium on Applied perception in graphics and visualization*, pages 151–151, New York, NY, USA, 2006. ACM Press.

[20] Fritz Ebner and Mark D. Fairchild. Development and testing of a color space (IPT) with improved hue uniformity. In *IS&T/SID Sixth Color Imaging Conference: Color Science, Systems and Applications*, pages 8–13, 1998. ISBN / ISSN: 0-89208-213-5.

[21] Garrett M. Johnson. *Measuring images: differences, quality and apperance*. PhD thesis, Rochester Institute of Technology, March 2003.

[22] Mark D. Fairchild. Considering the surround in device-independent color imaging. *Color Research and Application*, 20:352–363, 1995.

[23] Mark D. Fairchild. Modeling color appearance, spatial vision, and image quality imaging. *Color Image Science: Exploiting Digital Media*, pages 357–370, 2002.

[24] Garrett M. Johnson and Mark D. Fairchild. A top down description of S-CIELAB and CIEDE2000. *Color Research and Application*, 28(6):425–435, December 2003.

[25] Nathan Moroney, Mark D. Fairchild, Robert W.G. Hunt, Changjun Li, M. Ronnier Luo, and Todd Newman. The CIECAM02 color appearance model. *IST/SID 10th Color Imaging Conference*, pages 23–27, 2002.

[26] Mark D. Fairchild and Garrett M. Johnson. Meet iCAM: A next-generation color appearance model. *IS&T/SID Tenth Color Imaging Conference*, pages 33–38, 2002.

[27] Jan Morovic and Pei-Li Sun. Visual differences in colour reproduction and their colorimetric correlates. *IS&T Tenth Color Imaging Conference*, Scottsdale, Arizona:292–297, 2002.

[28] Pei-Li Sun and Jan Morovic. What differences do observers see in colour image reproduction experiments. In *CGIV 2002: The First European Conference on Colour Graphics, Imaging and Vision*, 2002.

[29] Scott Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In *Digital images and human vision*, pages 179–206. MIT Press, Cambridge, MA, USA, 1993.

[30] Andrew P. Bradley. A wavelet visible difference predictor. *IEEE Transactions on Image Processing*, 8:717–730, 1999.

[31] James A. Fenverda and Fabio Pellacini. Functional difference predictors (FDPs): measuring meaningful image differences. In *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1388 – 1392, November 2003.

[32] Laszlo Neumann, Kresimir Matkovic, and Werner Purgathofer. Perception based color image difference. Technical Report TR-186-2-97-21, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Favoritenstrasse 9-11/186, A-1040 Vienna, Austria, December 1997. human contact: technical-report@cg.tuwien.ac.at.

[33] Jason S. Babcock. Eye tracking observers during color image evaluation tasks. Master's thesis, Rochester Institute of Techonology, 2000.

[34] Jie Bai, Toshiya Nakaguchi, Norimichi Tsumura, and Yoichi Miyake. Evaluation of image corrected by retinex method based on S-CIELAB and gazing information. *IEICE trans. on Fundamentals of Electronics, Communications and Computer Sciences*, Vol.E89-A(11):2955–2961, 2006.

[35] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):802–816, 2006.

[36] David S. Wooding. Eye movements of large populations: II. deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, 34(4):518–528, November 2002.

[37] David S. Wooding. Fixation maps: quantifying eye-movement traces. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 31–36, New York, NY, USA, 2002. ACM Press.

[38] Norimichi Ukita, Tomohisa Ono, and Masatsugu Kidode. Region extraction of a gaze object using the gaze point and view image sequences. In *ICMI '05: Proceedings*

of the 7th international conference on Multimodal interfaces, pages 129–136, New York, NY, USA, 2005. ACM Press.

[39] Anthony Santella and Doug DeCarlo. Robust clustering of eye movement recordings for quantification of visual interest. In *ETRA '04: Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 27–34, New York, NY, USA, 2004. ACM Press.

[40] Anthony Santella and Doug DeCarlo. Visual interest and NPR: an evaluation and manifesto. In *NPAR '04: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering*, pages 71–150, New York, NY, USA, 2004. ACM Press.

[41] Kimiyoshi Miyata, Nasayuki Saito, Normichi Tsumura, Hideaki Haneishi, and Yoichi Miyake. Eye movement analysis and its application to evaluation of image quality. In *The Fifth Color Imaging Conference: Color Science, Systems, and Applications*, pages 116–119, 1997.

[42] Jonas Tillander. Design and implementation of a region of interest analysis system for eye-tracking studies. Master's thesis, STFI-Packforsk, 2004.

[43] Byoung Chul Ko and Jae-Yeal Nam. Automatic object-of-interest segmentation from natural images. *18th International Conference on Pattern Recognition (ICPR)*, 4(20-24):45–48, 2006.

[44] Jianping Fan, Xingquan Zhu, and Lide Wu. Automatic model-based semantic object extraction algorithm. *IEEE Transactions on Circuits and Systems for Video Technology*, 11:1073–1084, 2001.

[45] Wilfried Osberger, Neil Bergmann, and Anthony Maeder. An automatic image quality assessment technique incorporating higher level perceptual factors. In *Image Processing, ICIP 98*, volume 3, pages 414–418, 1998. 0-8186-8821-1.

[46] Claudio M. Privitera and Lawrence W. Stark. Algorithms for defining visual regions-of-interest: Comparision with eye fixations. *IEEE Transactions on pattern analysis and machine intelligence*, 22:970–981, 2000.

[47] Alfred L. Yarbus. *Eye Movements and Vision*. New York: Plenum Press, 1967. Original title: Rol' dvizenij glaz v processe zrenija.

[48] Sabira K. Mannan, Keith H. Ruddock, and David S. Wooding. The relationship between the locations of spatial features and those of fixations made during examination of briefly presented images. *Spatial Vision*, 10(3):165–188, 1996.

[49] Eriko Bando, Jon Y. Hardeberg, and David Connah. Can gamut mapping quality be predicted by color image difference formulae. In *Human Vision and Electronic Imaging X, ed. B. Rogowitz, T. Pappas, S. Daly, Proc. of SPIE - IST Electronic Imaging,SPIE*, volume 5666, pages 180 – 191, 2005.

[50] Wilfried Osberger and Anthony Maeder. Automatic identification of perceptually important regions in an image. In *International Conference on Pattern Recognition*, Brisbane, Australia, 1998.

[51] Joseph H. Goldberg, Mark J. Stimson, Marion Lewenstein, Neil Scott, and Anna M. Wichansky. Eye tracking in web search tasks: design implications. In *ETRA '02: Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 51–58, New York, NY, USA, 2002. ACM Press.

[52] Bing Pan, Helene A. Hembrooke, Geri K. Gay, Laura A. Granka, Matthew K. Feusner, and Jill K. Newman. The determinants of web page viewing behavior: an eye-tracking study. In *ETRA '04: Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 147–154, New York, NY, USA, 2004. ACM Press.

[53] Robert W. Reeder, Peter Pirolli, and Stuart K. Card. Webeyemapper and weblogger: tools for analyzing eye tracking data collected in web-use studies. In *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*, pages 19–20, New York, NY, USA, 2001. ACM Press.

[54] John D. McCarthy, M. Angela Sasse, and Jens Riegelsberger. Could I have the menu please? an eye tracking study of design conventions. http://www.cs.ucl.ac.uk/research/higherview/mccarthy_menu.pdf. Found 16/05/07.

[55] Laura Cowen, Linden J. Ball, and Judy Delin. An eye movement analysis of web-page usability. In *Proceedings of the annual HCI*, pages 317–335, 2002. 9/11/06: http://www.lauracowen.co.uk/.

[56] Jakob Nielsen. Eye-tracking research. 9/11/06: http://www.useit.com/eyetracking/.

[57] Sami Kukkonen. Exploring eye tracking in design evaluation. *Joining Forces*, pages 119 – 126, 2005.

[58] Kimmo Koivunen, Sami Kukkonen, Sami Lahtinen, Harri Rantala, and Selina Sharmin. Towards deeper understanding of how people perceive design in products. In *CADE2004 Web Proceedings of Computers in Art and Design Education Conference*, 2004.

[59] Tero Vuori, Maria Olkkonen, Monika Pölönen, Ari Siren, and Jukka Häkkinen. Can eye movements be quantitatively applied to image quality studies? In *NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction*, pages 335–338, New York, NY, USA, 2004. ACM Press.

[60] Benjamin Law, M. Stella Atkins, A. E. Kirkpatrick, and Alan J. Lomax. Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment. In *ETRA '04: Proceedings of the 2004 symposium on Eye tracking research & applications*, pages 41–48, New York, NY, USA, 2004. ACM Press.

[61] SMI (SensoMotoric Instruments). 02/04/07: http://www.smi.de.

[62] SMI (SensoMotoric Instruments). iViewX system manual, document version 1.05.09, 2005.

[63] Chengwu Cui. Comparision of two psychophysical methods for image color quality measurements: Paired comparision and rank order. In *IS&T/SID Eighth Color Imaging Conference*, pages 222–227, 2000.

[64] Stephan Brandt and Lawrence Stark. Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, 9:27–38, 1997.

[65] Amit Bhattacharya and James D. McGlothlin, editors. *Occupational Ergonomics: Theory and Applications*. Marcel Dekker, 1996.

[66] Hermann M. Burian and Gunter K. Von Noorden. *Burian-von Noorden's binocular vision and ocular motility : theory and management of strabismus*. St. Louis : Mosby, 1985.

[67] Eastman Kodak (Somadeepti N. Chengalur, Suzanne H. Rodgers, and Thomas E. Bernand), editors. *Kodak's Ergonomic Design for People at Work*. John Wiley and Sons, second edition, 2004. ISBN: 0-471-41863-3.

[68] Dennis R. Ankrum. Viewing angle and distance in computer workstations. 13/11/06: http://www.combo.com/ergo/vangle.htm.

[69] Sekonic. Sekonic light meters. http://www.sekonic.com.

[70] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical report, CIE TC8-08, 156:2004. ISBN: 3 901 906 26 6.

[71] Douglas A. Kerr. Apex - the additive system of photographic exposure. 16/3/07: http://doug.kerr.home.att.net/pumpkin/APEX.pdf, July 2006. Issue 5.

[72] Jeff Conrad. Exposure metering relating subject lighting to film exposure. 16/3/2007: http://www.largeformatphotography.info/articles/conrad-meter-cal.pdf, August 2003.

[73] Grafisk Assistans AB. Grafisk resursinfo.

[74] ISO. Graphic techonology - prepress digital echange. Technical report, ISO 12640-2, 2004. 1st edition.

[75] Mantas Malakauskas and Gediminas Montvilas. Panel testing for image quality. Technical report, Høgskolen i Gjøvik, 2003. Bachelor main project computer engeering. Supervisor: Ivar Farup.

[76] Jan Morovic. *To Develop a Universal Gamut Mapping Algorithm*. PhD thesis, University of Derby, 1998.

[77] Clare Porac and Stanley Coren. The dominant eye. *Psychological Bulletin*, 83(5):880–897, 1976.

[78] Joseph H. Goldberg and Anna M. Wichansky. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, chapter Eye tracking in Usability Evaluation: A Practitioner's Guide, pages 493–517. North-Holland, 2003.

[79] Heidi L. Roth, Andrea N. Lora, and Kenneth M. Heilman. Effects of monocular viewing and eye dominance on spatial attention. *Brain*, 125(9):2023–2035, September 2002.

[80] Agnieszika Bojko. Eye tracking in user experience testing: How to make the most of it. In *Proceedings of the UPA 2005 Conference*, 2005.

[81] H. Widdel. *Theoretical and Applied Aspects of Eye Movement Research*, chapter Operational problems in analysing eye movements, pages 22–29. Elsevier, New York, 1984.

[82] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau. A spatio-temporal model of the selective human visual attention. In *IEEE International Conference on Image Processing*, volume 3, pages 1188–91, 2005.

[83] Richard Godijn and Jan Theeuwes. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, chapter The Relationship Between Exogenous and Endogenous Saccades and Attention, pages 3–26. Elsevier, Amsterdam, 2003.

[84] Andrew P. Bradley and Fred W.M. Stentiford. Visual attention for region of interest coding in JPEG 2000. *Journal of Visual Communication & Image representation*, 14:232 – 250, 2003.

[85] Scott Daly, Kristine Matthews, and Jordi Ribas-Corbera. Face-based visually-optimized image sequence coding. In *International Conference on Image Processing, ICIP 98.*, volume 3, pages 443–447, 1998. ISBN: 0-8186-8821-1.

[86] John C. Russ. *The image processing handbook*. CRC Press, 2nd ed. edition, 1995.

[87] Ralph Radach, Stefanie Lemmer, Christian Vorstius, Dieter Heller, and Karina Radach. *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, chapter Eye movements in the processing of print advertisements, pages 609–632. Elsevier, Amsterdam, 2003.

[88] Gunnar G. Løvås. *Statistikk - for universiteter og høgskoler*. Universitetsforlaget, 1999.

[89] The MathWorks Inc. Documentation image processing toolbox: rgb2gray. 28/05/07: http://www.mathworks.com.

A Questionnaire

Observer number:

Age:

Have you studied color science?
[ ] Yes [ ] No

Have you participated in eye tracking experiments before?
[ ] Yes [ ] No

Have you participated in psychophysical experiments before?
[ ] Yes [ ] No

Have you seen any of the images used in this experiment before?

(a)

(b)

(c)

(d)

Image a: [ ] Yes [ ] No
Image b: [ ] Yes [ ] No
Image c: [ ] Yes [ ] No
Image d: [ ] Yes [ ] No

Did you recognize the man on the second image(image b)?
[ ] Yes [ ] No

Please mark the regions important for your choice

Please mark the regions important for your choice

Please mark the regions important for your choice

Please mark the regions important for your choice

B QuickEval 2.0

This appendix contain supplementary information about the pair comparison program QuickEval 2.0.

B.1 Communicating with iViewX

This code is an excerpt from QuickEval 2.0, it only shows an example of communicating with iViewX.

```
...
 try{
        DatagramSocket socket;
        DatagramPacket packet,packetpic;
        InetAddress address;
        int pnum = Integer.parseInt(portn);
        address = InetAddress.getByName(ipcom);
        socket = new DatagramSocket();
        // ET_REC to begin recoring (can be replaced with desired code)
        String mess = "ET_REC"+'\n';
        //send MSG message to iViewX
        String messpic = "ET_REM \"TRG: START RECORDING"+"\""+'\n';
        byte message[] = mess.getBytes();
        byte messagepic[] = messpic.getBytes();
        packet = new DatagramPacket(message, message.length,
        address, pnum);
        packetpic = new DatagramPacket(messagepic,
        messagepic.length, address, pnum);
        socket.send(packet);
        socket.send(packetpic);
 }
      catch(IOException io){}
   }
...
```

B.2 QuickEval image configuration

This section includes the image configuration from QuickEval and the order they were put in.

B.2.1 Scene1

GIRL_LM3.PNG
GIRL_LM5.PNG
GIRL_LP3.PNG
GIRL_LP5.PNG

GIRL_REGION_BACK_LM3.PNG
GIRL_REGION_BACK_LP3.PNG
GIRL_REGION_FRONT_LM3.PNG
GIRL_REGION_FRONT_LP3.PNG

B.2.2 Scene2

TORE_LM3.PNG
TORE_LM5.PNG
TORE_LP3.PNG
TORE_LP5.PNG
TORE_REGION_BACK_LM3.PNG
TORE_REGION_BACK_LP3.PNG
TORE_REGION_FRONT_LM3.PNG
TORE_REGION_FRONT_LP3.PNG

B.2.3 Scene3

JP_LM3.PNG
JP_LM5.PNG
JP_LP3.PNG
JP_LP5.PNG
JP_REGION_STRIPE2_LM3.PNG
JP_REGION_STRIPE2_LP3.PNG
JP_REGION_STRIPE_LM3.PNG
JP_REGION_STRIPE_LP3.PNG

B.2.4 Scene4

CARTOON_LM3.PNG
CARTOON_LM5.PNG
CARTOON_LP3.PNG
CARTOON_LP5.PNG
CARTOON_REGION_SIGN_LM3.PNG
CARTOON_REGION_SIGN_LP3.PNG
CARTOON_REGION_TSHIRT_LM3.PNG
CARTOON_REGION_TSHIRT_LP3.PNG

B.3 Raw data from QuickEval

This is the raw data from which the z-scores have been calculated. The order of the images (horizontal and vertical) correspond to the order in the configuration file in Section B.2.

0	12	27	16	30	33	30	21
38	0	40	22	47	45	44	34
23	10	0	10	34	34	28	27
34	28	40	0	39	43	36	40
20	3	16	11	0	23	25	22
17	5	16	7	27	0	22	14
20	6	22	14	25	28	0	21
29	16	23	10	28	36	29	0

Table 31: Raw z-score matrix for scene 1.

0	11	19	15	34	31	22	28
39	0	32	22	42	41	38	37
31	18	0	18	35	29	33	25
35	28	32	0	38	37	41	37
16	8	15	12	0	23	21	18
19	9	21	13	27	0	13	22
28	12	17	9	29	37	0	28
22	13	25	13	32	28	22	0

Table 32: Raw z-score matrix for scene 2.

0	6	25	9	28	31	14	13
44	0	36	23	35	40	22	27
25	14	0	12	31	31	17	13
41	27	38	0	42	41	18	18
22	15	19	8	0	28	10	9
19	10	19	9	22	0	11	10
36	28	33	32	40	39	0	16
37	23	37	32	41	40	34	0

Table 33: Raw z-score matrix for scene 3.

0	8	18	13	35	35	37	34
42	0	38	25	45	47	47	47
32	12	0	10	34	37	41	43
37	25	40	0	45	43	47	45
15	5	16	5	0	29	27	28
15	3	13	7	21	0	26	25
13	3	9	3	23	24	0	23
16	3	7	5	22	25	27	0

Table 34: Raw z-score matrix for scene 4.

C How do we look at images

This appendix section contain extra information for the chapter "How do we look at images".

C.1 Difference between maps

All maps for the different groups are found in this section.

C.1.1 Observer maps for different groups

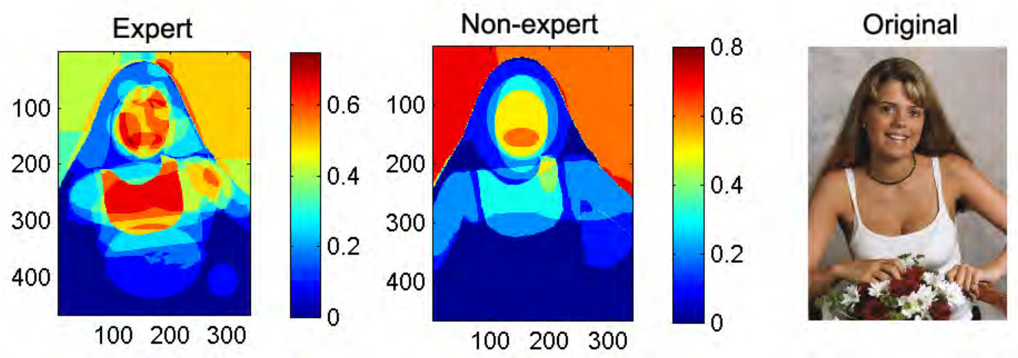


Figure 118: Observers map from experts and non-experts in Girl scene.

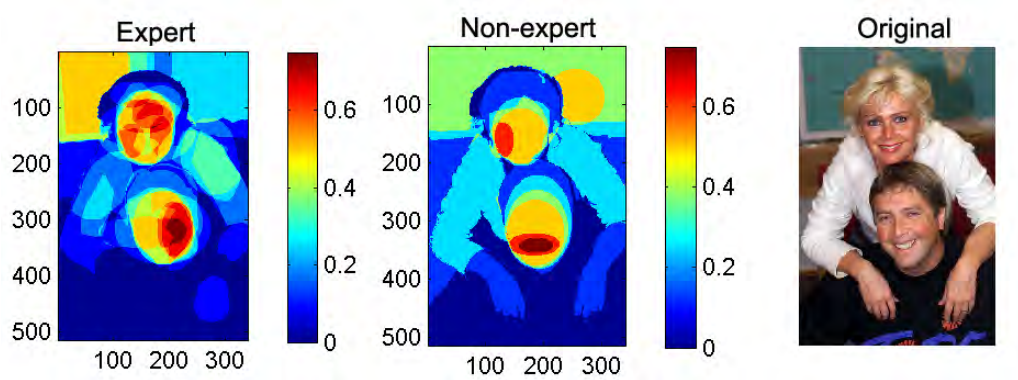


Figure 119: Observers map from experts and non-experts in Tore scene.

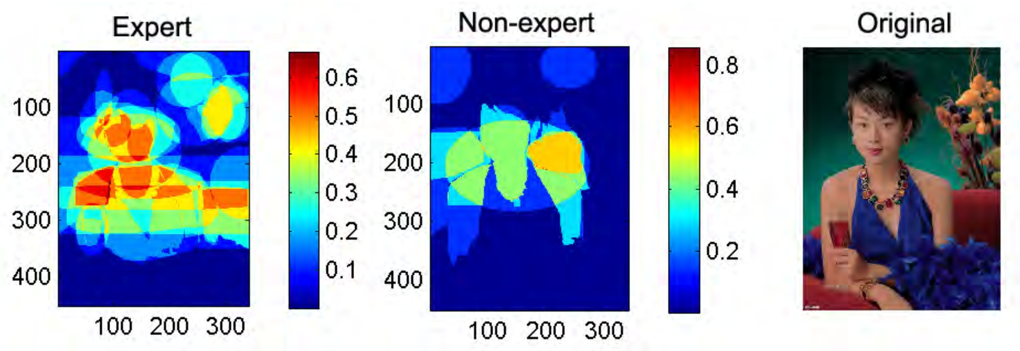


Figure 120: Observers map from experts and non-experts in JP scene.

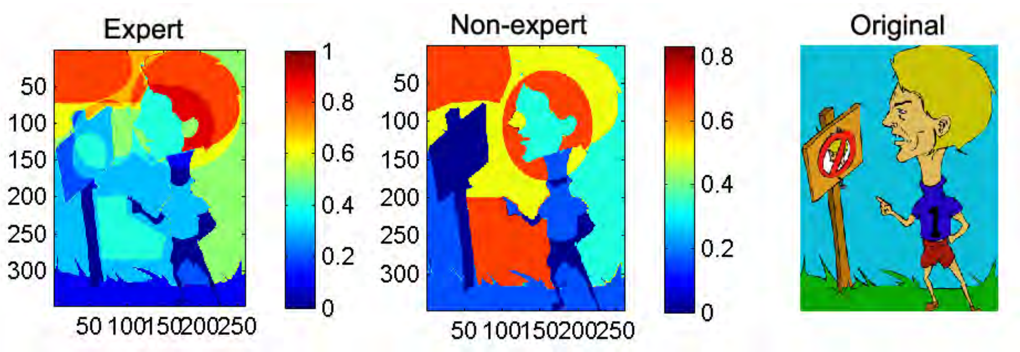


Figure 121: Observers map from experts and non-experts in Cartoon scene.

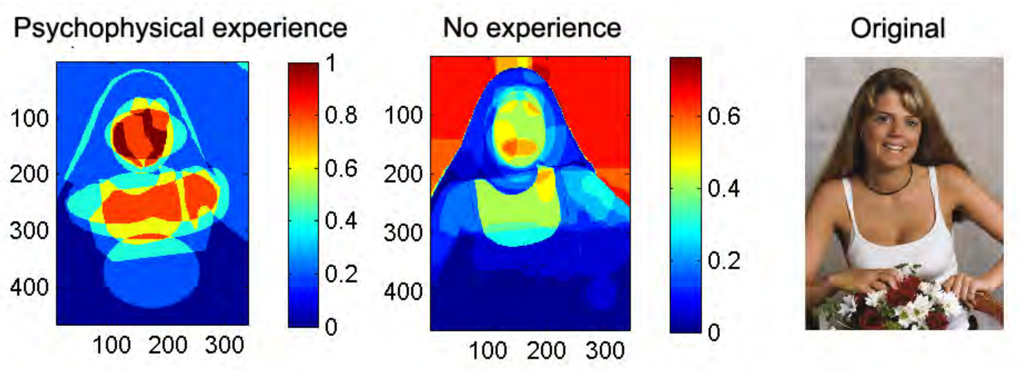


Figure 122: Observers map from psychophysical experience and no experience in Girl scene.

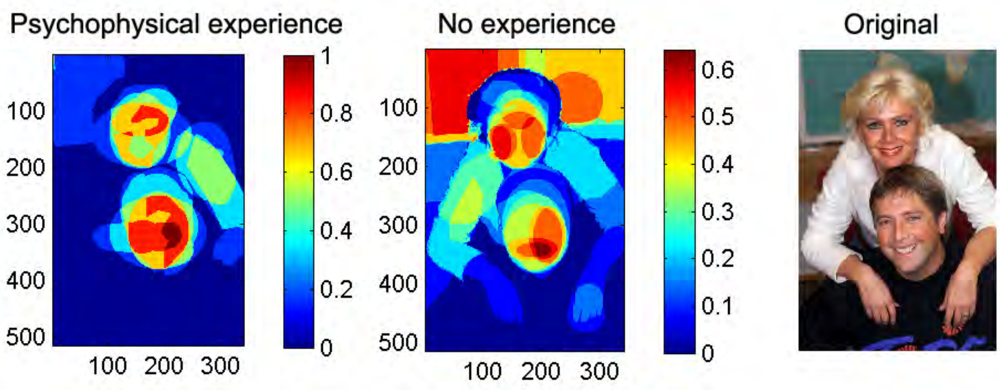


Figure 123: Observers map from psychophysical experience and no experience in Tore scene.

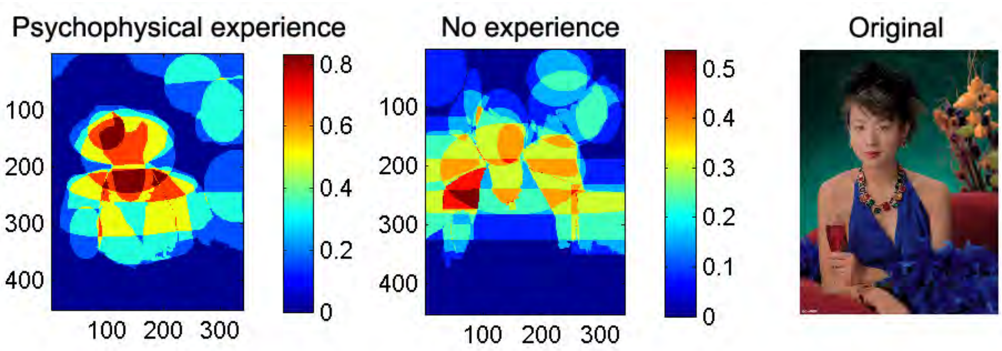


Figure 124: Observers map from psychophysical experience and no experience in JP scene.

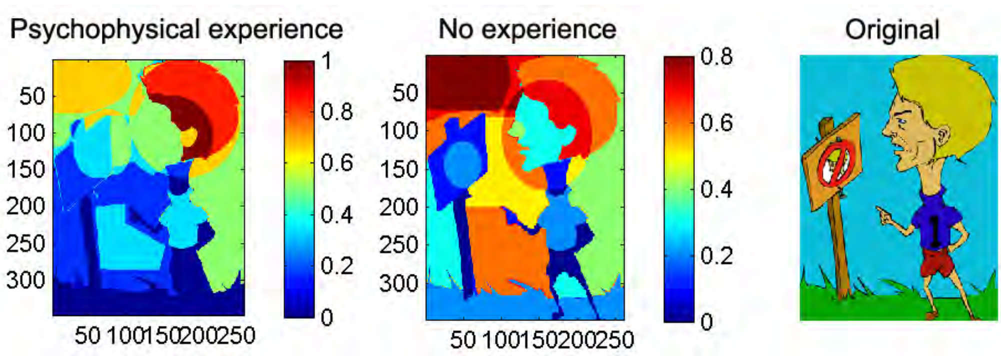


Figure 125: Observers map from psychophysical experience and no experience in Cartoon scene.

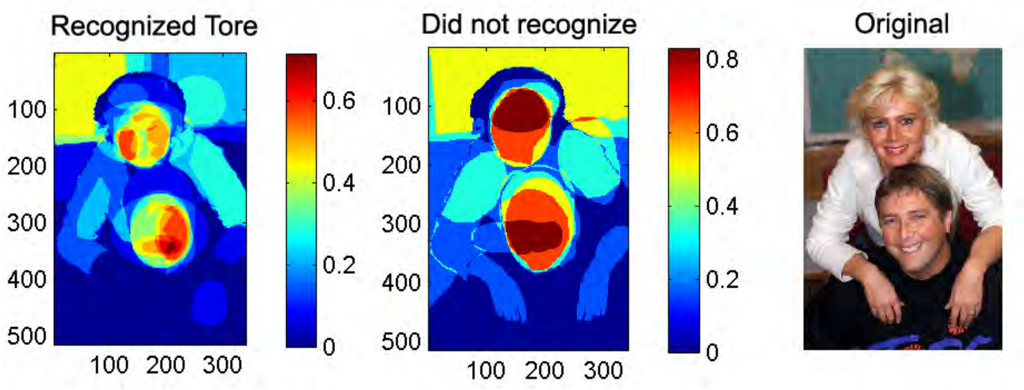


Figure 126: Observers map from recognized Tore and did not recognize Tore.

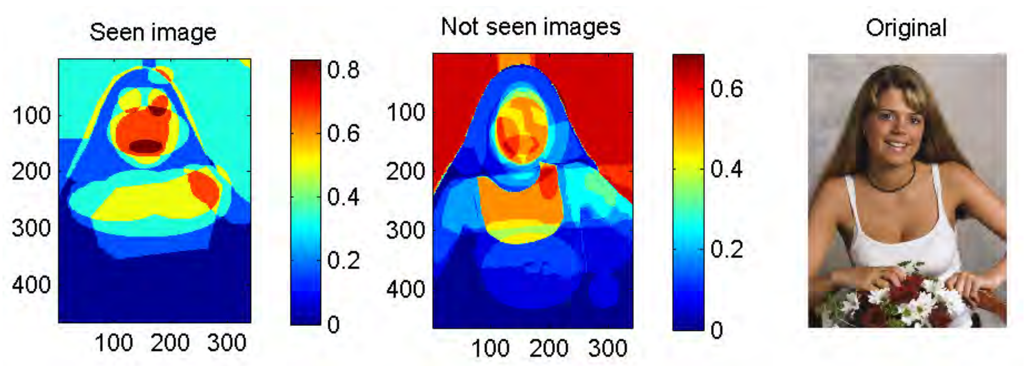


Figure 127: Observers map from seen image and not seen image in Girl scene.

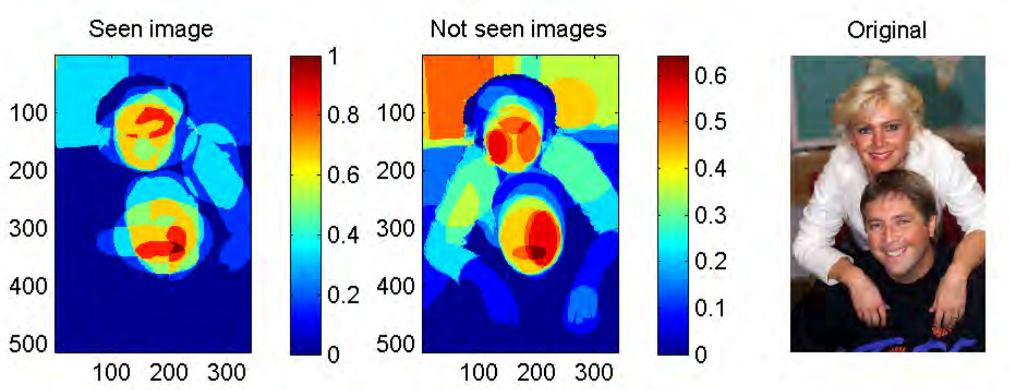


Figure 128: Observers map from seen image and not seen image in Tore scene.

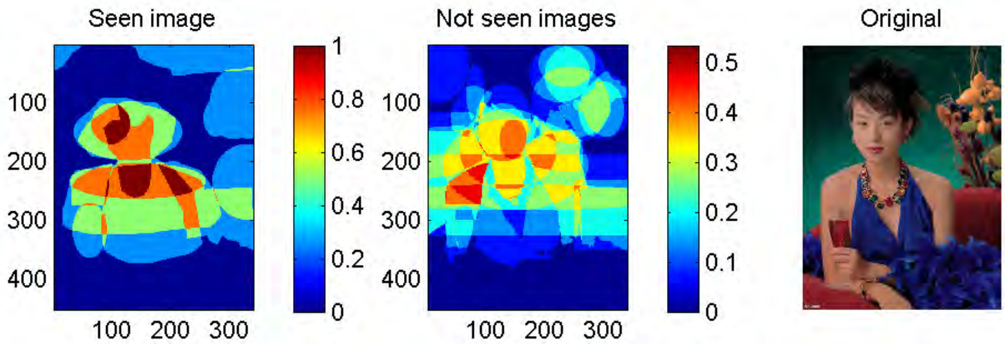


Figure 129: Observers map from seen image and not seen image in JP scene.

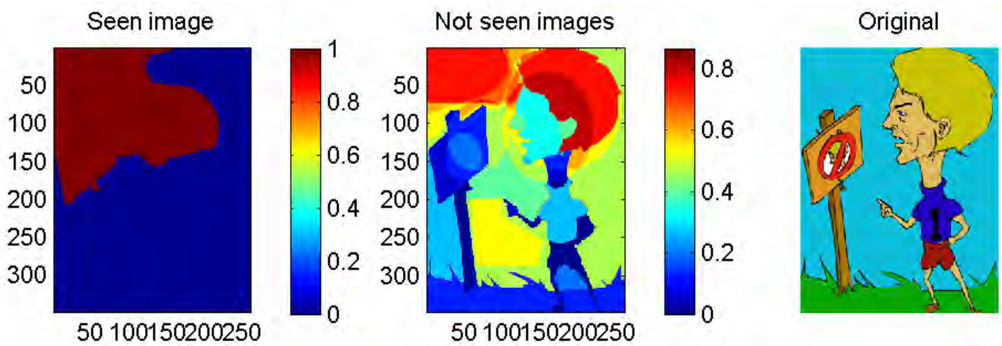


Figure 130: Observers map from seen image and not seen image in Cartoon scene.

C.1.2 Binary eye tracking maps for different groups

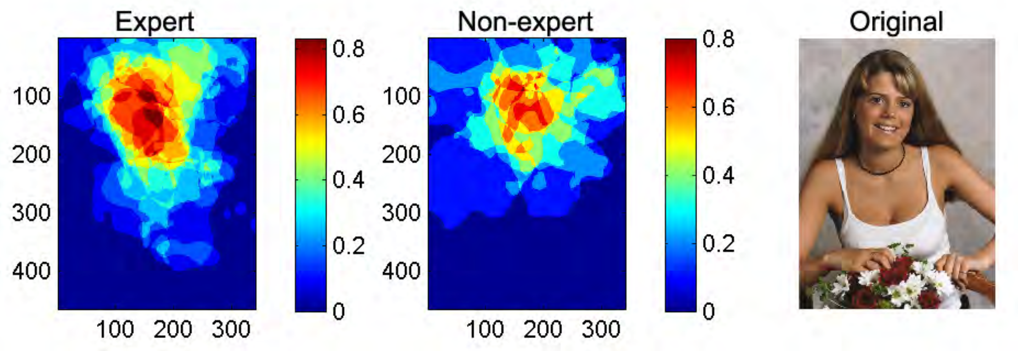


Figure 131: Eye tracker map from experts and non-experts in Girl scene.

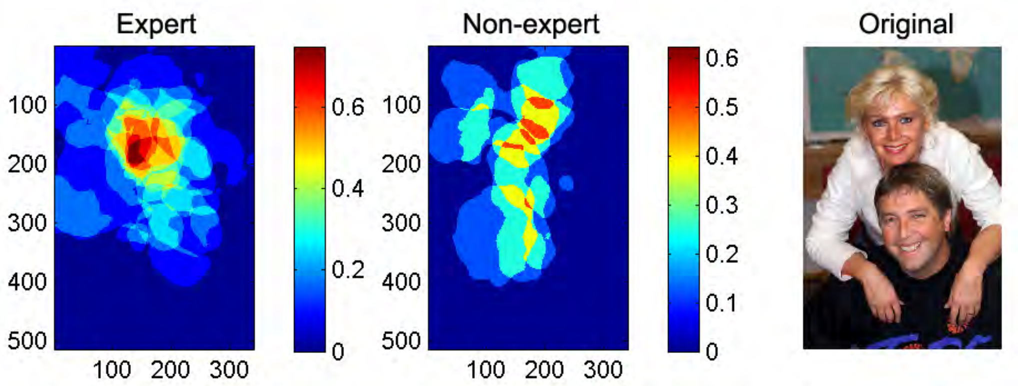


Figure 132: Eye tracker map from experts and non-experts in Tore scene.

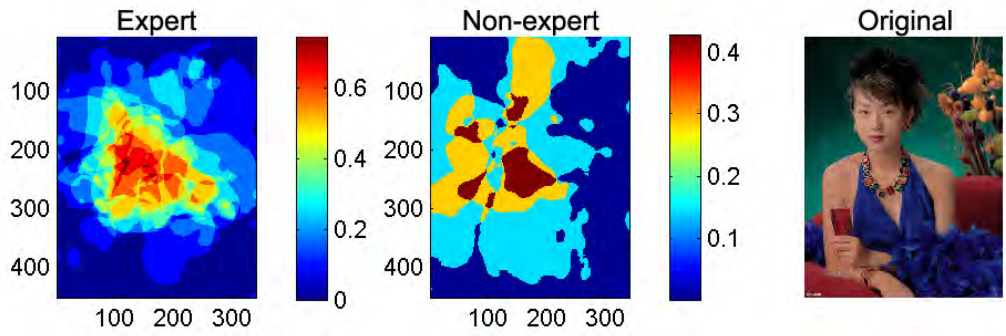


Figure 133: Eye tracker map from experts and non-experts in JP scene.

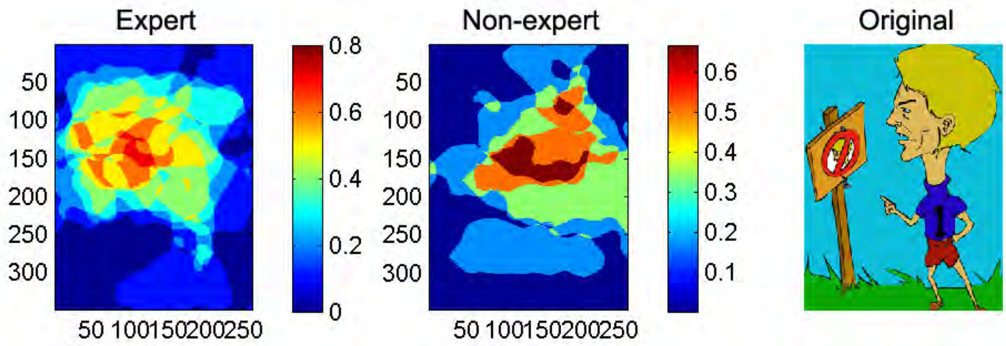


Figure 134: Eye tracker map from experts and non-experts in Cartoon scene.

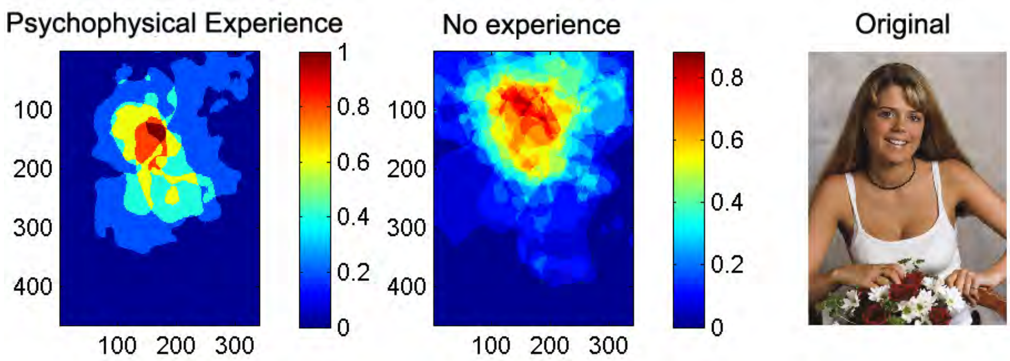


Figure 135: Eye tracker map from psychophysical experience and no experience in Girl scene.

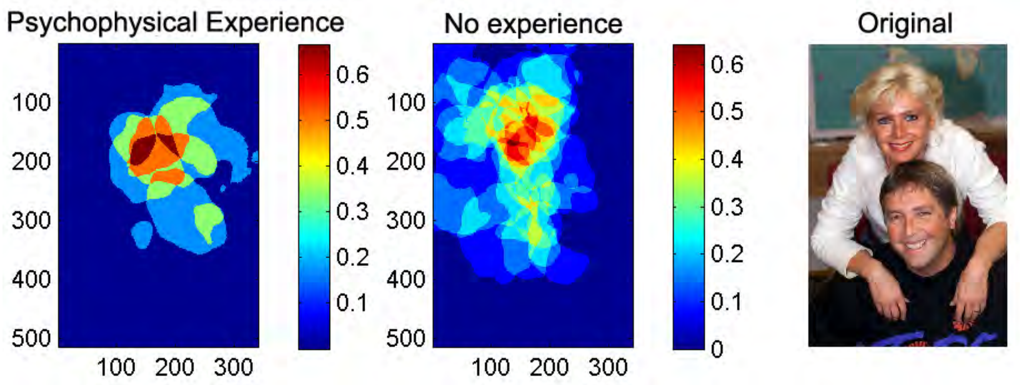


Figure 136: Eye tracker map from psychophysical experience and no experience in Tore scene.

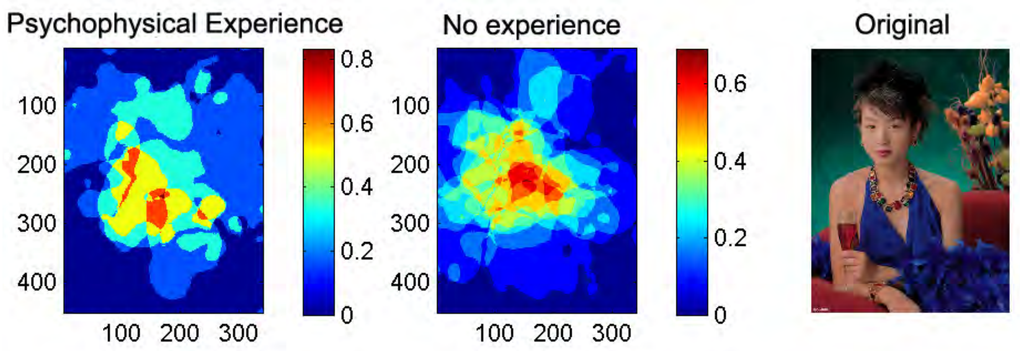


Figure 137: Eye tracker map from psychophysical experience and no experience in JP scene.

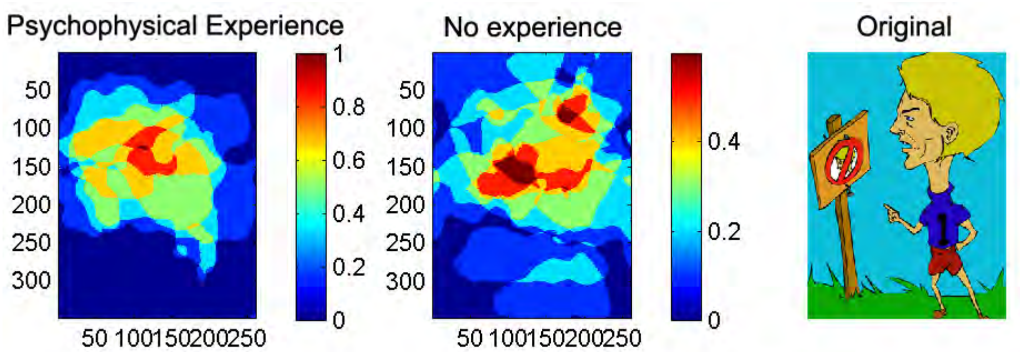


Figure 138: Eye tracker map from psychophysical experience and no experience in Cartoon scene.

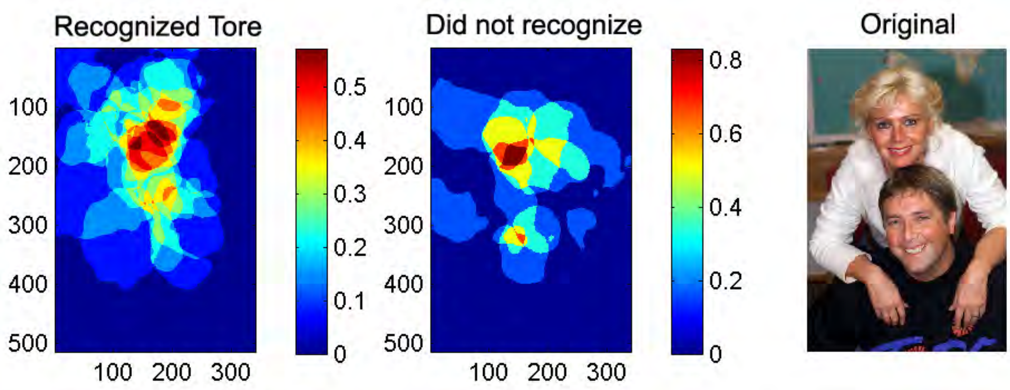


Figure 139: Eye tracker map from recognized Tore and did not recognize Tore.

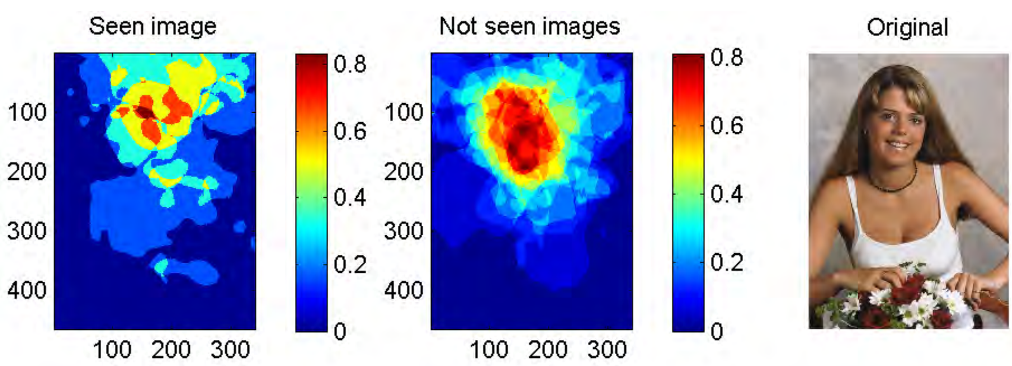


Figure 140: Eye tracker map from seen image and not seen image in Girl scene.

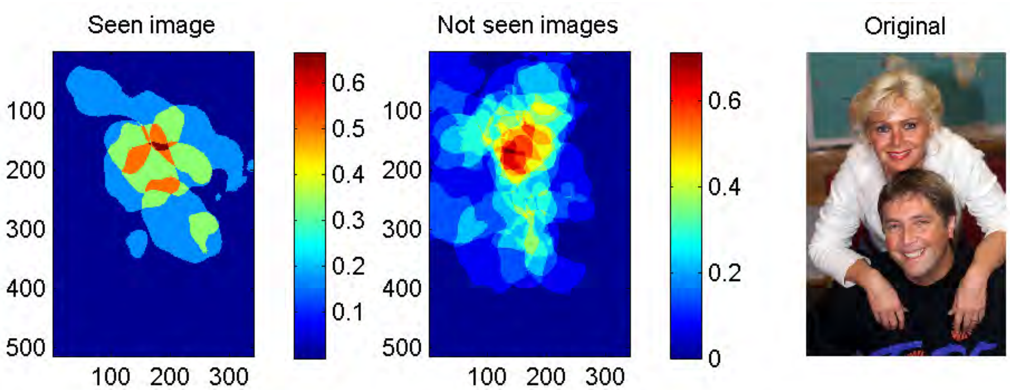


Figure 141: Eye tracker map from seen image and not seen image in Tore scene.

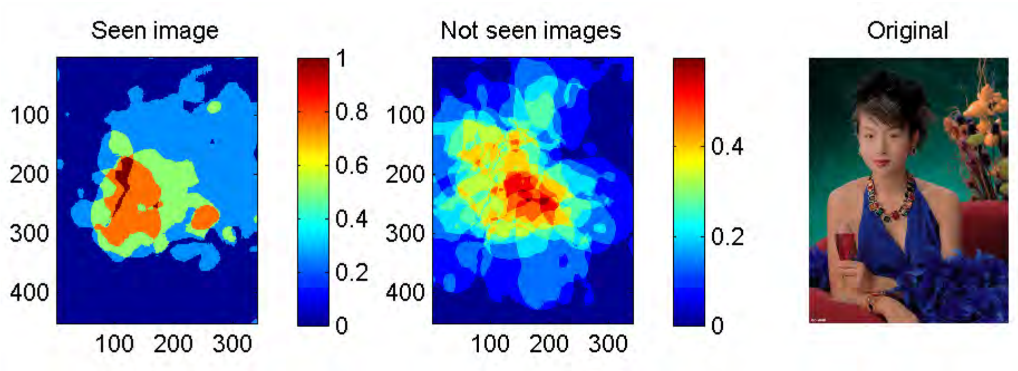


Figure 142: Eye tracker map from seen image and not seen image in JP scene.

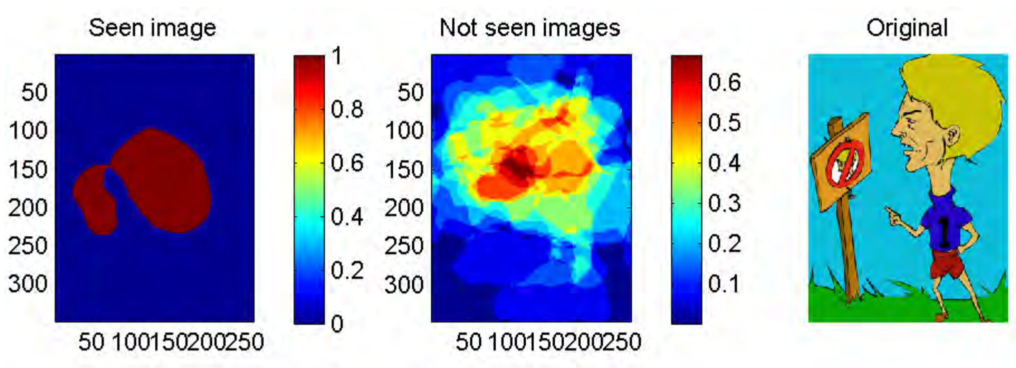


Figure 143: Eye tracker map from seen image and not seen image in Cartoon scene.

C.1.3 Observer stated regions and binary eye tracking map

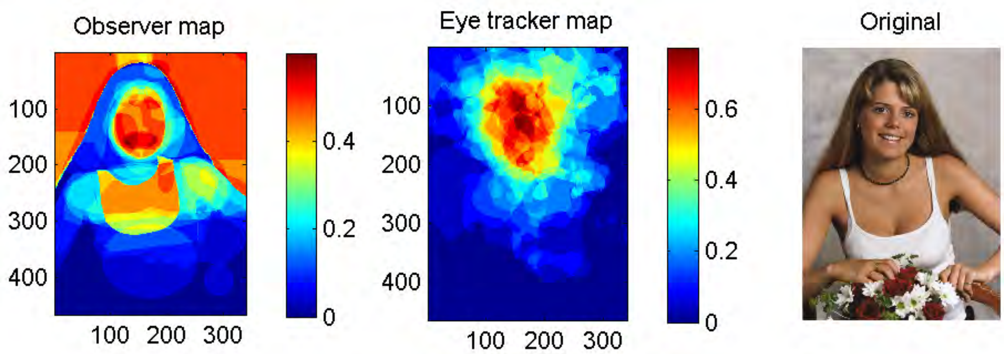


Figure 144: Overall observer and eye tracking map for Girl scene.

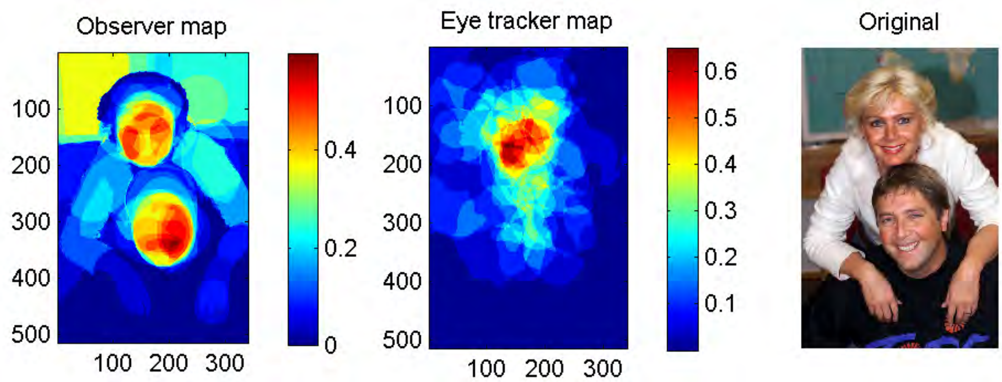


Figure 145: Overall observer and eye tracking map for Tore scene.

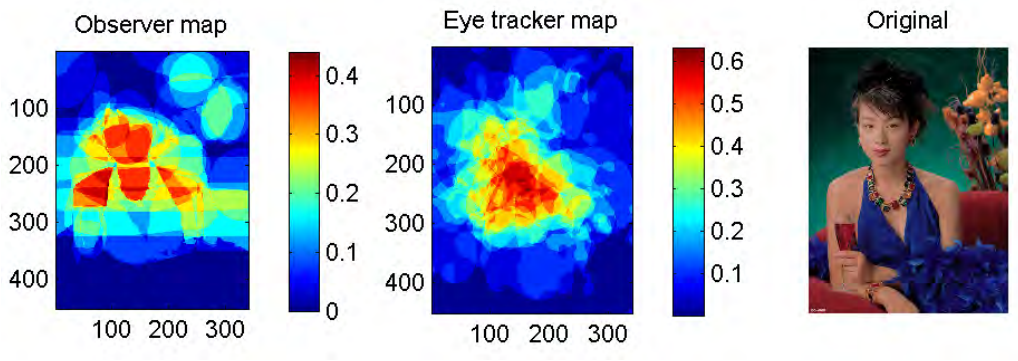


Figure 146: Overall observer and eye tracking map for JP scene.

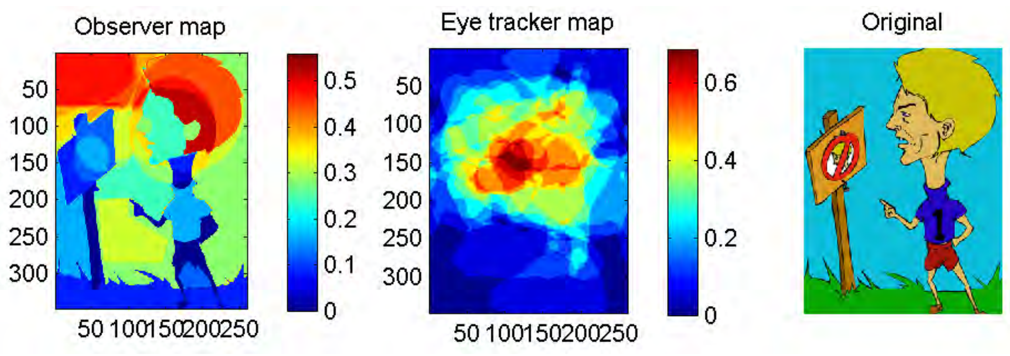
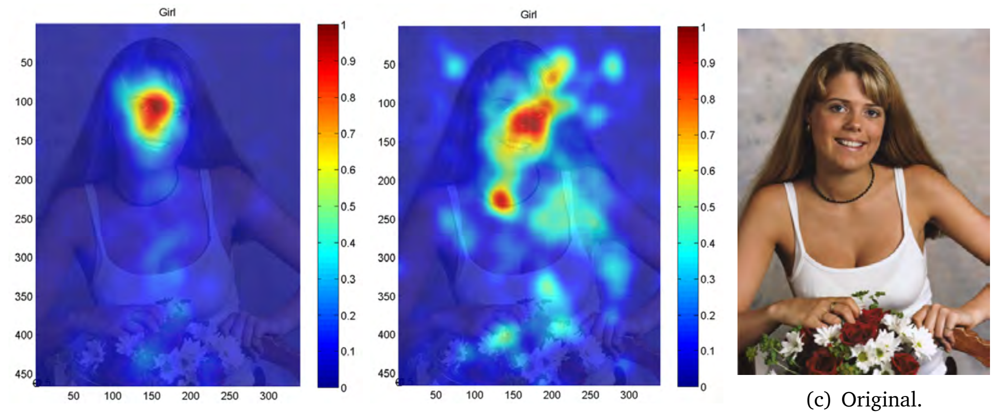


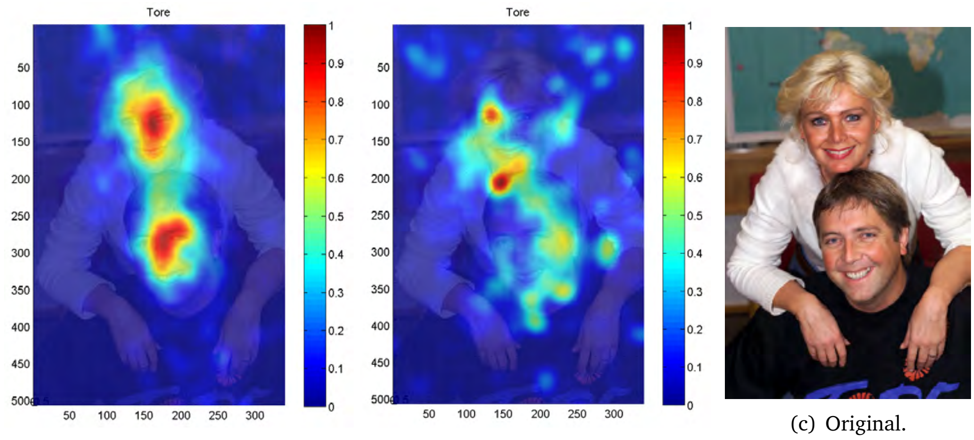
Figure 147: Overall observer and eye tracking map for Cartoon scene.

C.1.4 Freeview and gaze marking



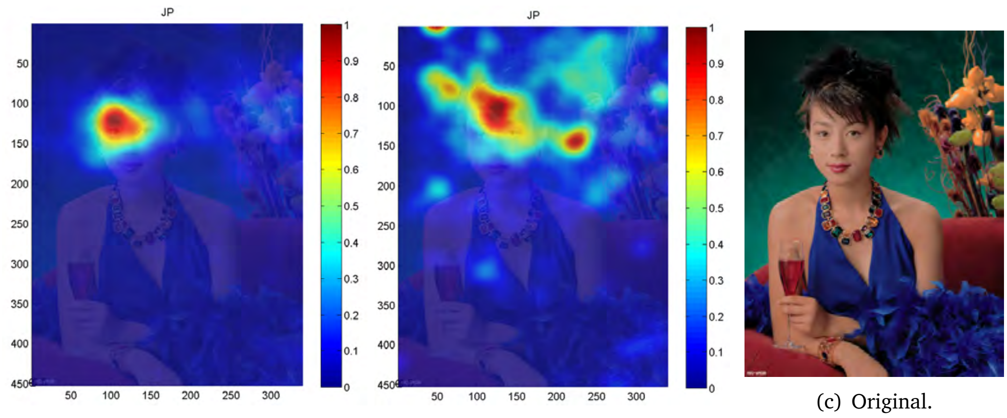
(a) Freeview. (b) Gaze marking. (c) Original.

Figure 148: Maps from freeview and gaze marking for Girl scene.



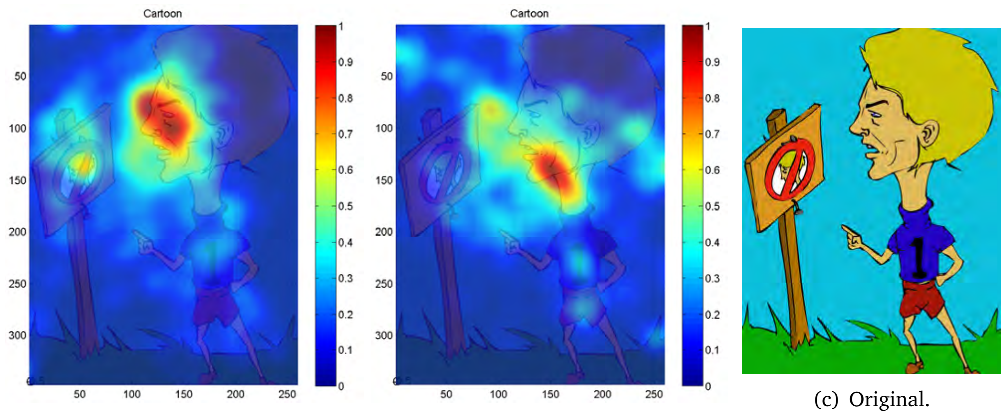
(a) Freeview. (b) Gaze marking. (c) Original.

Figure 149: Maps from freeview and gaze marking for Tore scene.



(a) Freeview. (b) Gaze marking. (c) Original.

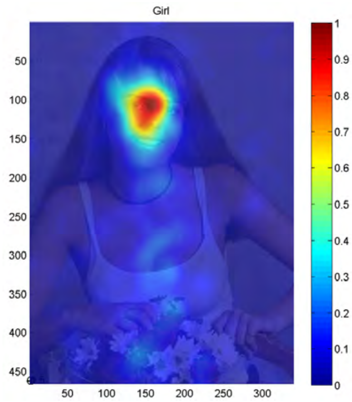
Figure 150: Maps from freeview and gaze marking for JP scene.



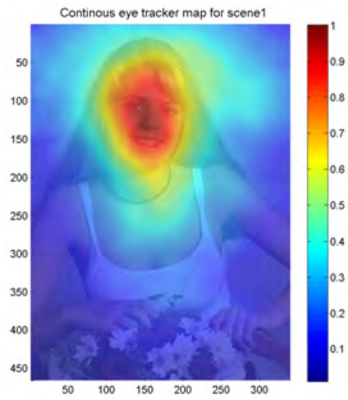
(a) Freeview. (b) Gaze marking. (c) Original.

Figure 151: Maps from freeview and gaze marking for Cartoon scene.

C.1.5 Freeview and continuous eye tracker maps



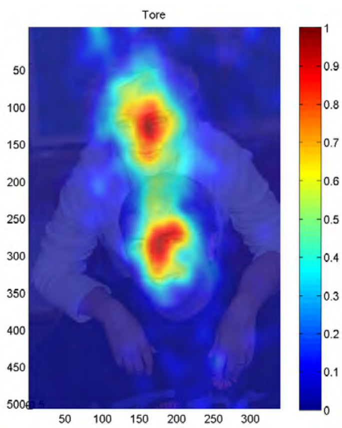
(a) Freeview.



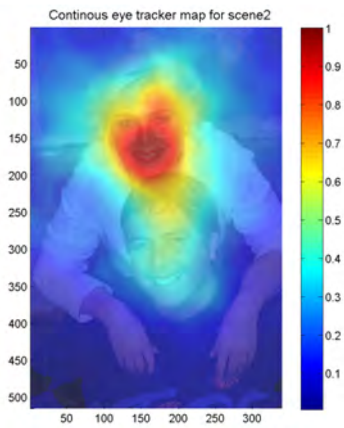
(b) Pair comparison.

(c) Original.

Figure 152: Maps from freeview and pair comparison in Girl scene.



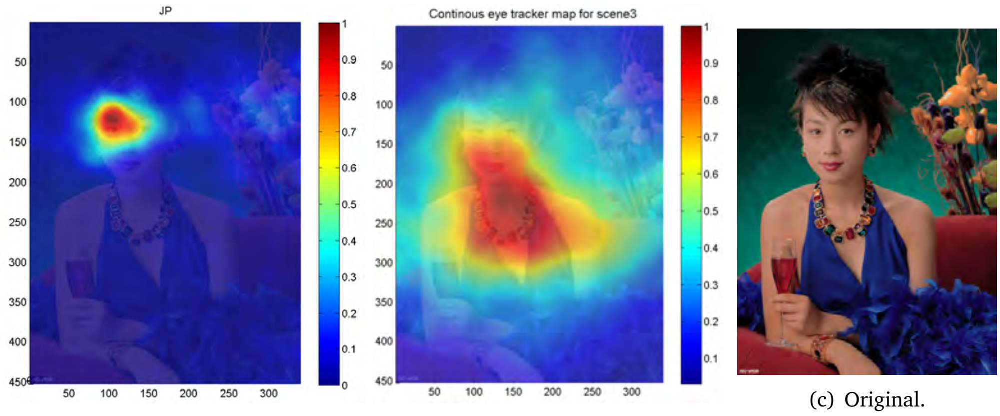
(a) Freeview.



(b) Pair comparison.

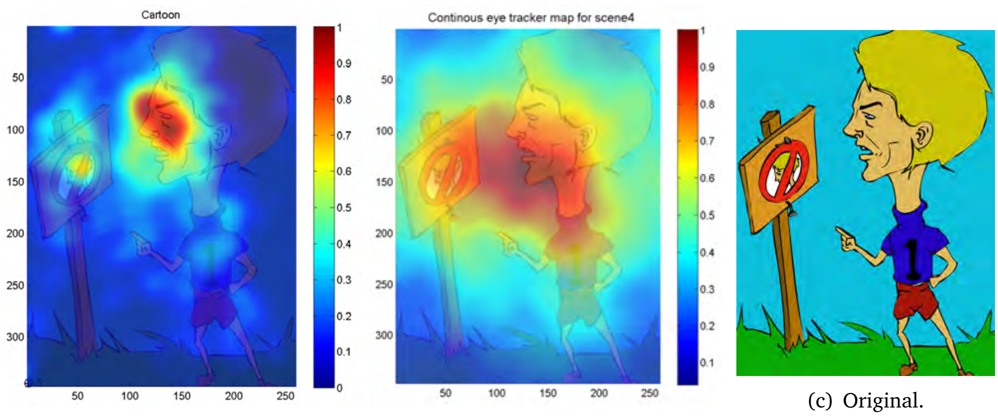
(c) Original.

Figure 153: Maps from freeview and pair comparison in Tore scene.



(a) Freeview. (b) Pair comparison. (c) Original.

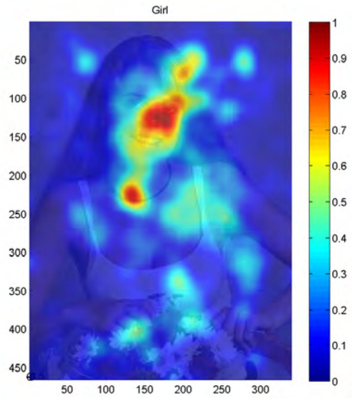
Figure 154: Maps from freeview and pair comparison in JP scene.



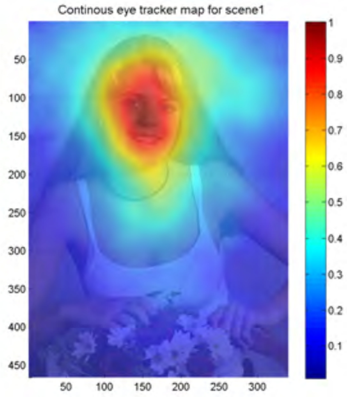
(a) Freeview. (b) Pair comparison. (c) Original.

Figure 155: Maps from freeview and pair comparison in Cartoon scene.

C.1.6 Gaze marking and eye tracker continuous



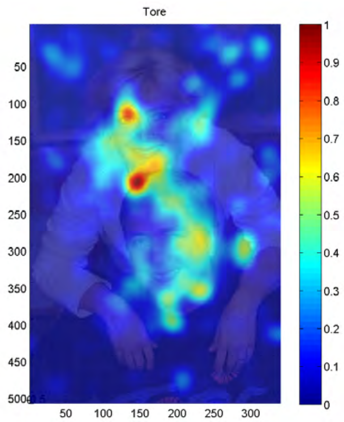
(a) Gaze marking.



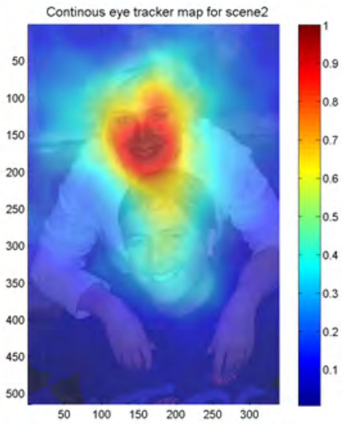
(b) Pair comparison.

(c) Original.

Figure 156: Maps from gaze marking and pair comparison in Girl scene.



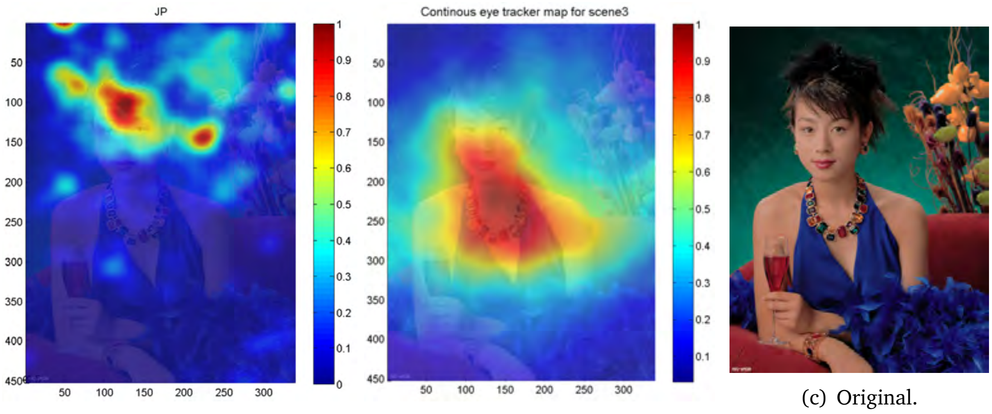
(a) Gaze marking.



(b) Pair comparison.

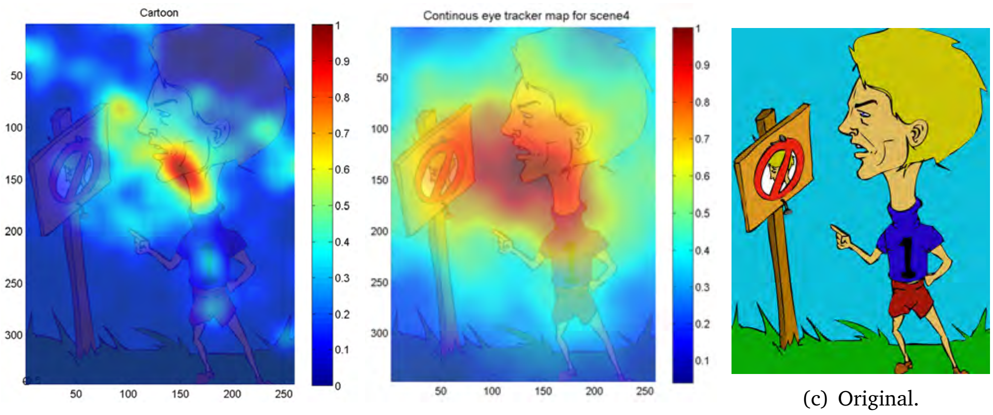
(c) Original.

Figure 157: Maps from gaze marking and pair comparison in Tore scene.



(a) Gaze marking. (b) Pair comparison. (c) Original.

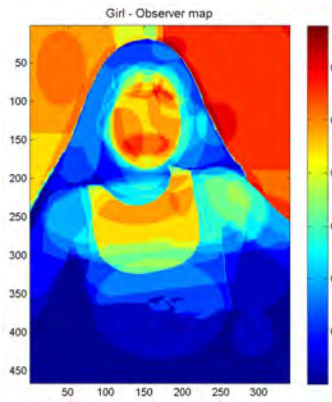
Figure 158: Maps from gaze marking and pair comparison in JP scene.



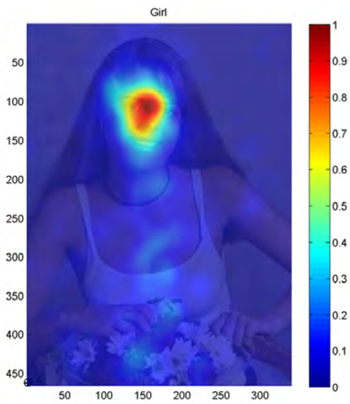
(a) Gaze marking. (b) Pair comparison. (c) Original.

Figure 159: Maps from gaze marking and pair comparison in Cartoon scene.

C.1.7 Observer regions and freeview



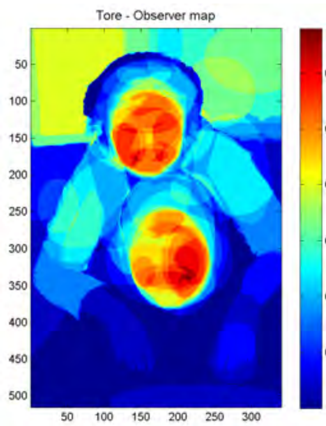
(a) Observer regions.



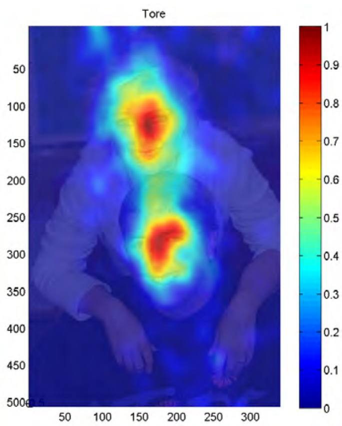
(b) Freeview.

(c) Original.

Figure 160: Maps from observer regions and freeview in Girl scene.



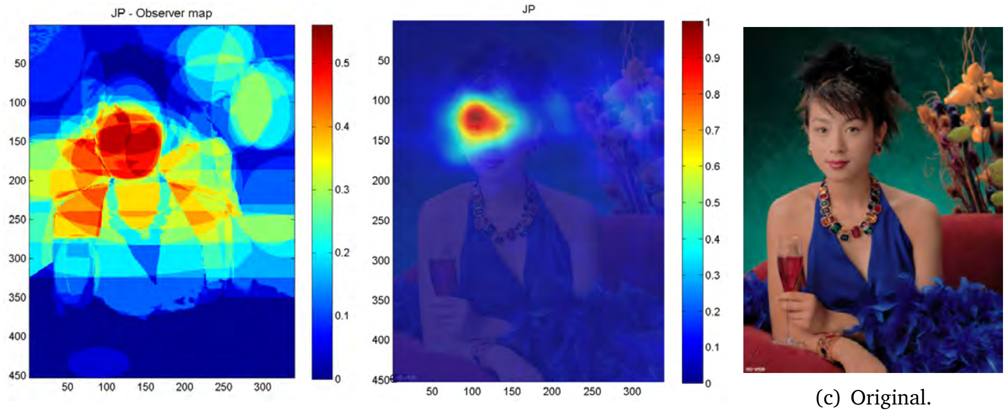
(a) Observer regions.



(b) Freeview.

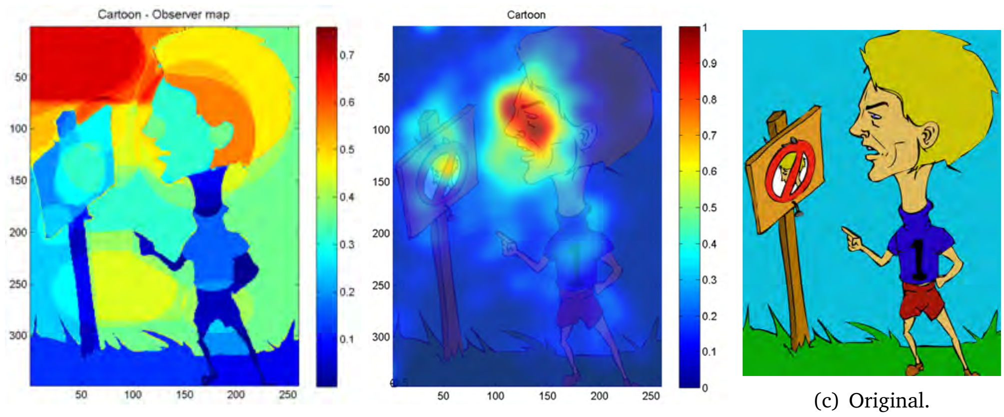
(c) Original.

Figure 161: Maps from observer regions and freeview in Tore scene.



(a) Observer regions. (b) Freeview. (c) Original.

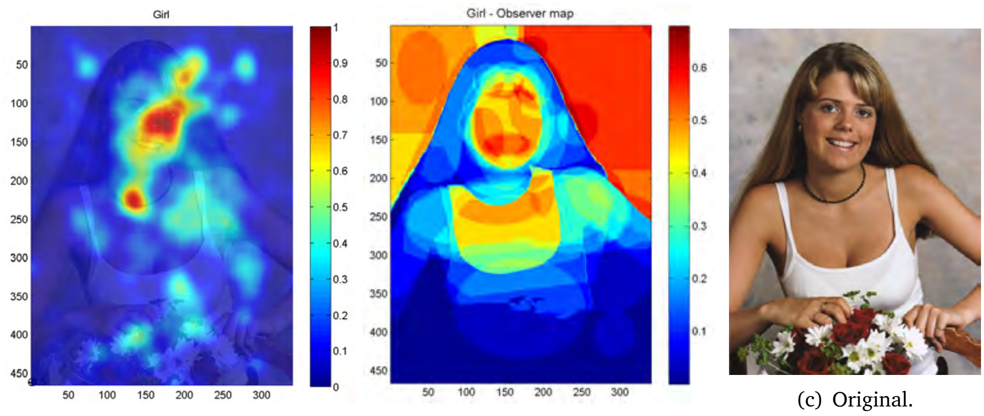
Figure 162: Maps from observer regions and freeview in JP scene.



(a) Observer regions. (b) Freeview. (c) Original.

Figure 163: Maps from observer regions and freeview in Cartoon scene.

C.1.8 Observer regions and gaze marking

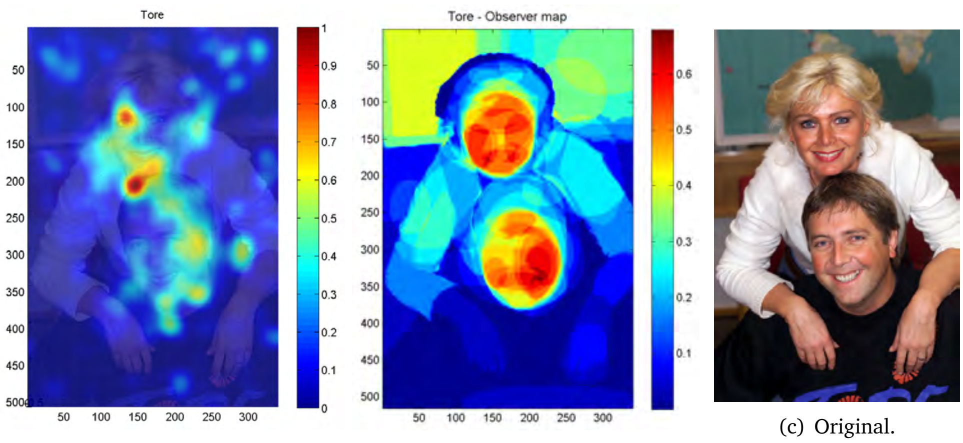


(a) Gaze marking.

(b) Observer regions.

(c) Original.

Figure 164: Map from gaze marking and observer marked regions in Girl scene.

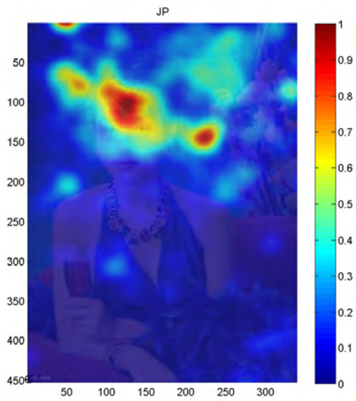


(a) Gaze marking.

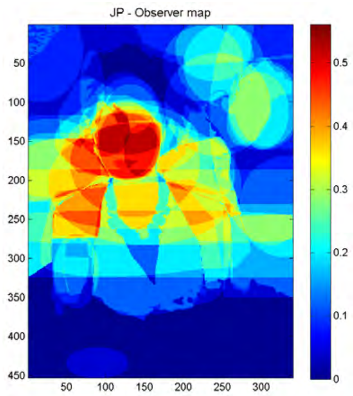
(b) Observer regions.

(c) Original.

Figure 165: Map from gaze marking and observer marked regions in Tore scene.



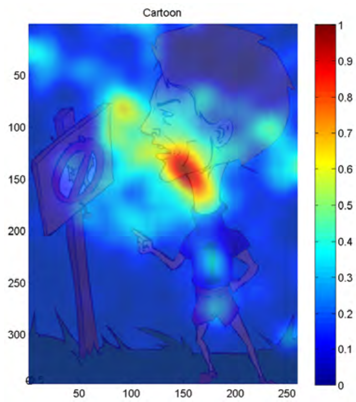
(a) Gaze marking.



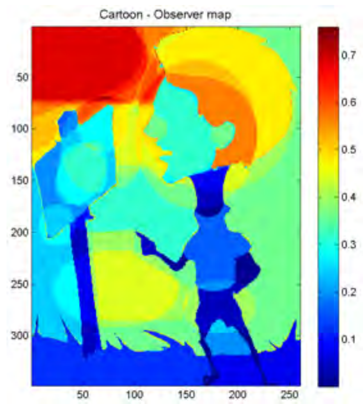
(b) Observer regions.

(c) Original.

Figure 166: Map from gaze marking and observer marked regions in JP scene.



(a) Gaze marking.



(b) Observer regions.

(c) Original.

Figure 167: Map from gaze marking and observer marked regions in Cartoon scene.

C.1.9 Differences between experts and non-experts

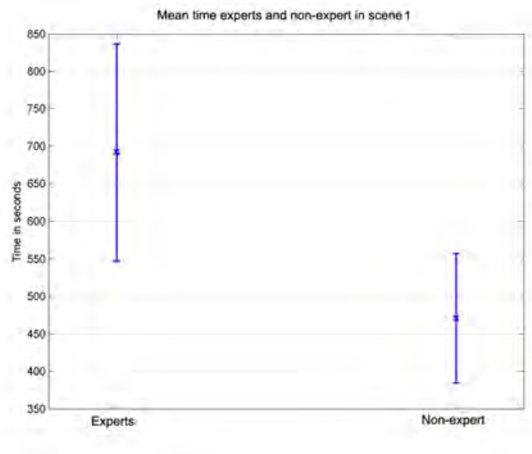


Figure 168: Time difference between experts and non-experts for scene 1.

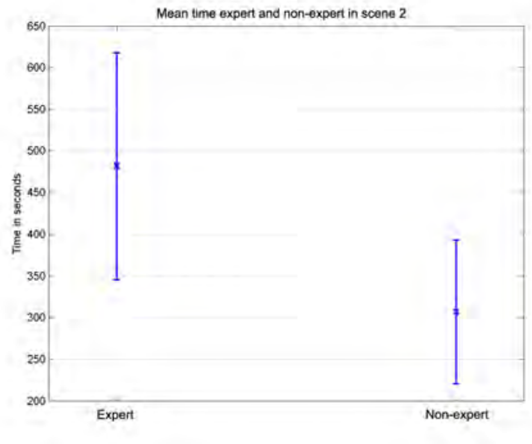


Figure 169: Time difference between experts and non-experts for scene 2.

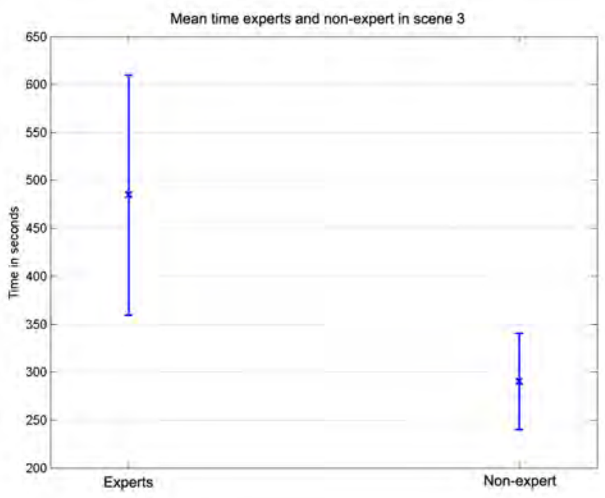


Figure 170: Time difference between experts and non-experts for scene 3.

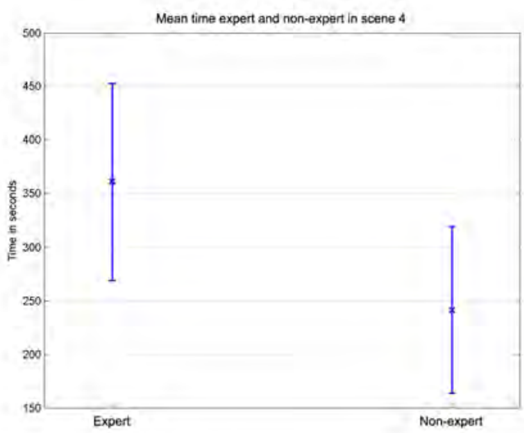
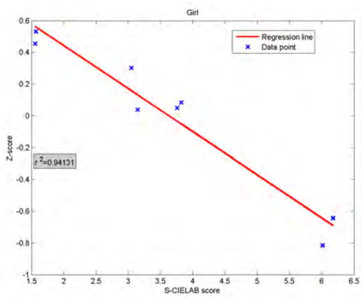


Figure 171: Time difference between experts and non-experts for scene 4.

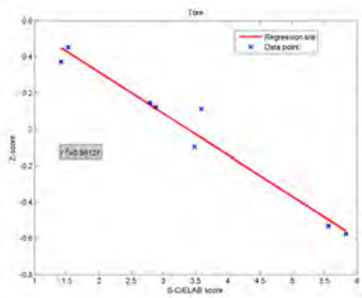
D Image difference metrics

This appendix contain supplementary information and data to the chapter "Image difference metrics".

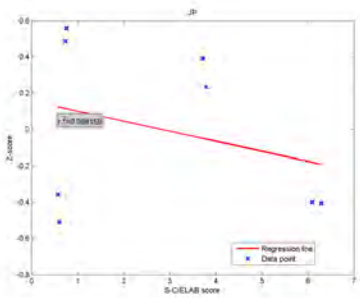
D.1 Image difference metrics results



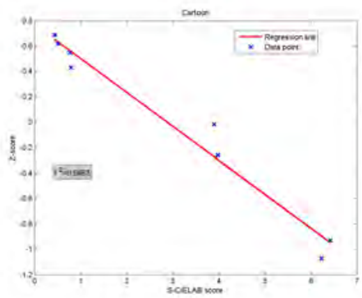
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 172: S-CIELAB regression plot for each scene.

Scene	Mean
CARTOON_LM3	3.897
CARTOON_LM5	6.228
CARTOON_LP3	3.982
CARTOON_LP5	6.418
CARTOON_REGION_SIGN_LM3	0.792
CARTOON_REGION_SIGN_LP3	0.772
CARTOON_REGION_TSHIRT_LM3	0.437
CARTOON_REGION_TSHIRT_LP3	0.510
GIRL_LM3	3.760
GIRL_LM5	6.012
GIRL_LP3	3.820
GIRL_LP5	6.174
GIRL_REGION_BACK_LM3	1.558
GIRL_REGION_BACK_LP3	1.569
GIRL_REGION_FRONT_LM3	3.049
GIRL_REGION_FRONT_LP3	3.145
JP_LM3	3.716
JP_LM5	6.287
JP_LP3	3.794
JP_LP5	6.097
JP_REGION_STRIPE2_LM3	0.736
JP_REGION_STRIPE2_LP3	0.761
JP_REGION_STRIPE_LM3	0.576
JP_REGION_STRIPE_LP3	0.608
TORE_LM3	3.586
TORE_LM5	5.829
TORE_LP3	3.482
TORE_LP5	5.556
TORE_REGION_BACK_LM3	1.522
TORE_REGION_BACK_LP3	1.411
TORE_REGION_FRONT_LM3	2.881
TORE_REGION_FRONT_LP3	2.793

Table 35: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3	0.418
CARTOON_LM5	0.633
CARTOON_LP3	0.807
CARTOON_LP5	1.312
CARTOON_REGION_SIGN_LM3	0.199
CARTOON_REGION_SIGN_LP3	0.192
CARTOON_REGION_TSHIRT_LM3	0.072
CARTOON_REGION_TSHIRT_LP3	0.079
GIRL_LM3	0.659
GIRL_LM5	0.999
GIRL_LP3	0.958
GIRL_LP5	1.508
GIRL_REGION_BACK_LM3	0.473
GIRL_REGION_BACK_LP3	0.483
GIRL_REGION_FRONT_LM3	0.830
GIRL_REGION_FRONT_LP3	0.973
JP_LM3	0.712
JP_LM5	1.028
JP_LP3	0.796
JP_LP5	1.035
JP_REGION_STRIPE2_LM3	0.338
JP_REGION_STRIPE2_LP3	0.535
JP_REGION_STRIPE_LM3	0.211
JP_REGION_STRIPE_LP3	0.218
TORE_LM3	0.764
TORE_LM5	1.042
TORE_LP3	1.340
TORE_LP5	1.922
TORE_REGION_BACK_LM3	0.499
TORE_REGION_BACK_LP3	0.448
TORE_REGION_FRONT_LM3	0.837
TORE_REGION_FRONT_LP3	1.280

Table 36: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3	3.766
CARTOON_LM5	5.898
CARTOON_LP3	3.365
CARTOON_LP5	5.585
CARTOON_REGION_SIGN_LM3	0.359
CARTOON_REGION_SIGN_LP3	0.337
CARTOON_REGION_TSHIRT_LM3	0.107
CARTOON_REGION_TSHIRT_LP3	0.110
GIRL_LM3	3.137
GIRL_LM5	5.092
GIRL_LP3	3.142
GIRL_LP5	5.099
GIRL_REGION_BACK_LM3	1.002
GIRL_REGION_BACK_LP3	1.002
GIRL_REGION_FRONT_LM3	2.135
GIRL_REGION_FRONT_LP3	2.139
JP_LM3	3.201
JP_LM5	5.454
JP_LP3	3.151
JP_LP5	5.108
JP_REGION_STRIPE2_LM3	0.476
JP_REGION_STRIPE2_LP3	0.472
JP_REGION_STRIPE_LM3	0.270
JP_REGION_STRIPE_LP3	0.272
TORE_LM3	3.059
TORE_LM5	4.957
TORE_LP3	3.117
TORE_LP5	5.029
TORE_REGION_BACK_LM3	0.987
TORE_REGION_BACK_LP3	0.868
TORE_REGION_FRONT_LM3	2.072
TORE_REGION_FRONT_LP3	2.249

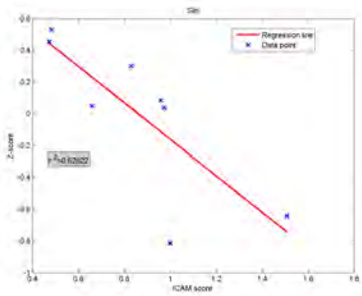
Table 37: ΔE^*_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3	0.991
CARTOON_LM5	0.985
CARTOON_LP3	0.992
CARTOON_LP5	0.988
CARTOON_REGION_SIGN_LM3	0.993
CARTOON_REGION_SIGN_LP3	0.998
CARTOON_REGION_TSHIRT_LM3	0.995
CARTOON_REGION_TSHIRT_LP3	0.994
GIRL_LM3	0.994
GIRL_LM5	0.980
GIRL_LP3	0.997
GIRL_LP5	0.991
GIRL_REGION_BACK_LM3	0.995
GIRL_REGION_BACK_LP3	0.996
GIRL_REGION_FRONT_LM3	0.991
GIRL_REGION_FRONT_LP3	0.993
JP_LM3	0.877
JP_LM5	0.801
JP_LP3	0.913
JP_LP5	0.894
JP_REGION_STRIPE2_LM3	0.984
JP_REGION_STRIPE2_LP3	0.992
JP_REGION_STRIPE_LM3	0.990
JP_REGION_STRIPE_LP3	0.989
TORE_LM3	0.842
TORE_LM5	0.828
TORE_LP3	0.905
TORE_LP5	0.869
TORE_REGION_BACK_LM3	0.987
TORE_REGION_BACK_LP3	0.995
TORE_REGION_FRONT_LM3	0.850
TORE_REGION_FRONT_LP3	0.906

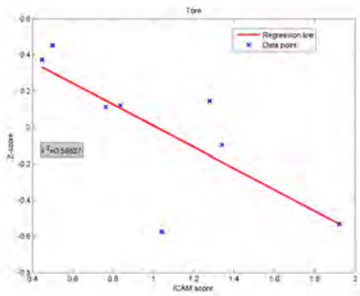
Table 38: SSIM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3	3.073
CARTOON_LM5	8.088
CARTOON_LP3	3.119
CARTOON_LP5	8.169
CARTOON_REGION_SIGN_LM3	0.059
CARTOON_REGION_SIGN_LP3	0.056
CARTOON_REGION_TSHIRT_LM3	0.024
CARTOON_REGION_TSHIRT_LP3	0.031
GIRL_LM3	2.376
GIRL_LM5	6.293
GIRL_LP3	2.379
GIRL_LP5	6.286
GIRL_REGION_BACK_LM3	0.251
GIRL_REGION_BACK_LP3	0.253
GIRL_REGION_FRONT_LM3	1.463
GIRL_REGION_FRONT_LP3	1.467
JP_LM3	2.841
JP_LM5	8.258
JP_LP3	2.862
JP_LP5	6.815
JP_REGION_STRIPE2_LM3	0.058
JP_REGION_STRIPE2_LP3	0.050
JP_REGION_STRIPE_LM3	0.045
JP_REGION_STRIPE_LP3	0.044
TORE_LM3	2.883
TORE_LM5	7.160
TORE_LP3	2.858
TORE_LP5	7.068
TORE_REGION_BACK_LM3	0.306
TORE_REGION_BACK_LP3	0.239
TORE_REGION_FRONT_LM3	2.135
TORE_REGION_FRONT_LP3	2.288

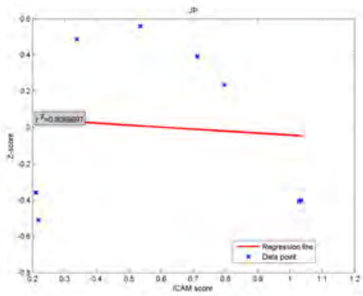
Table 39: Hue angle score for all scenes and reproductions.



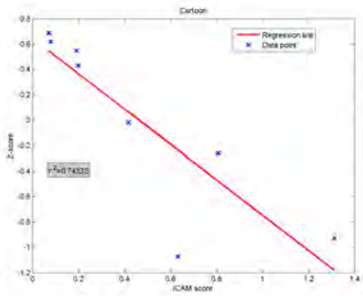
(a) Girl.



(b) Tore.

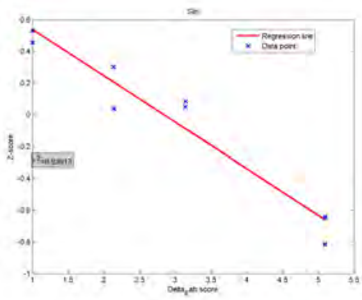


(c) JP.

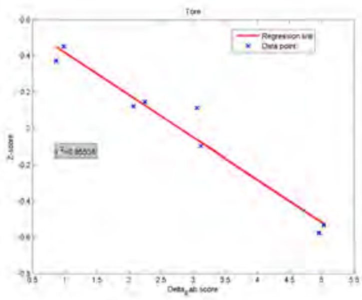


(d) Cartoon.

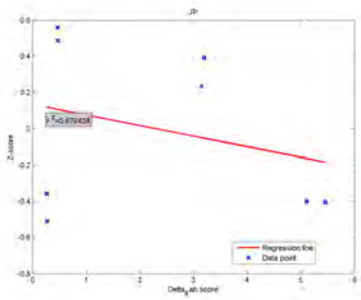
Figure 173: iCAM regression plot for each scene.



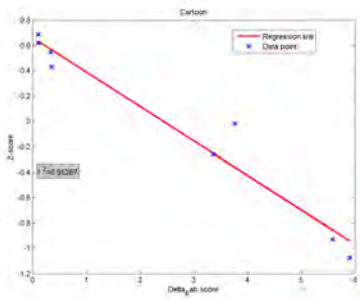
(a) Girl.



(b) Tore.

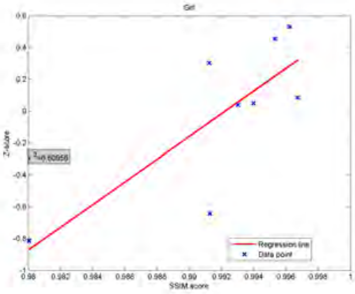


(c) JP.

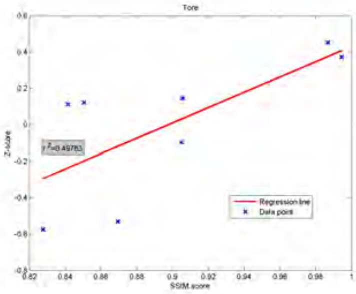


(d) Cartoon.

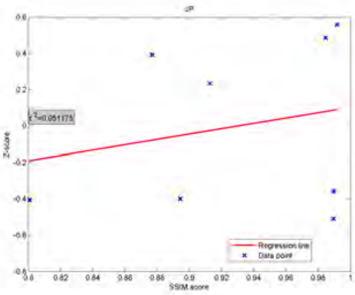
Figure 174: ΔE_{ab} regression plot for each scene.



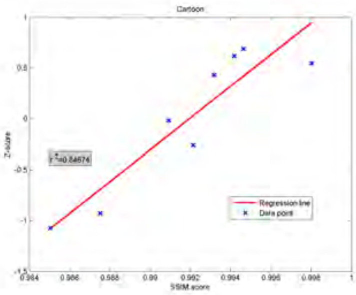
(a) Girl.



(b) Tore.

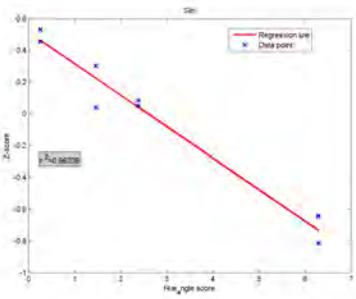


(c) JP.

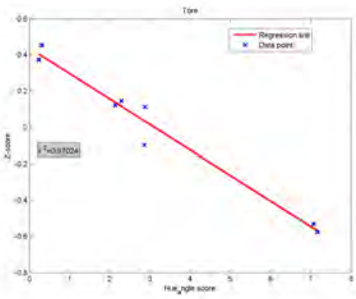


(d) Cartoon.

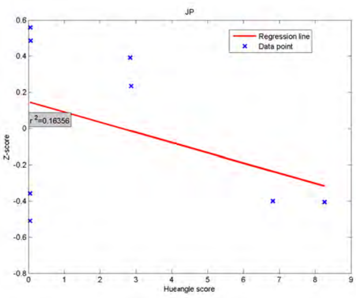
Figure 175: SSIM regression plot for each scene.



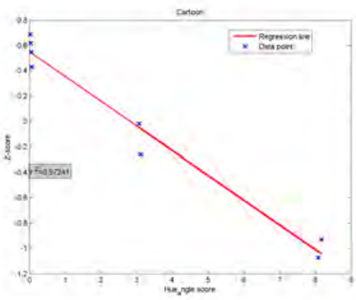
(a) Girl.



(b) Tore.

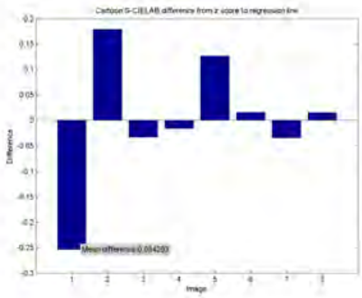


(c) JP.

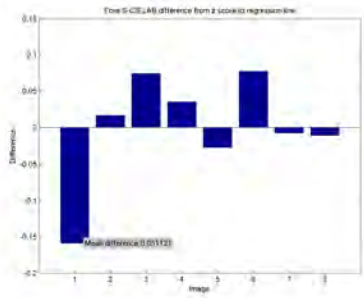


(d) Cartoon.

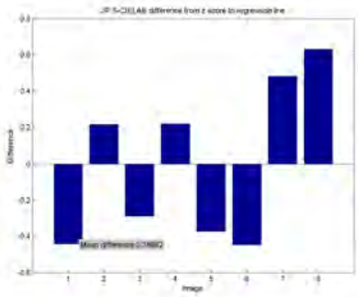
Figure 176: Hue angle regression plot for each scene.



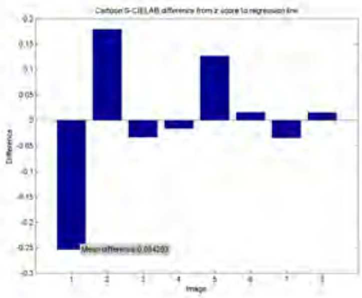
(a) Girl.



(b) Tore.

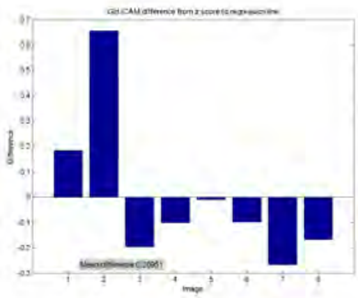


(c) JP.

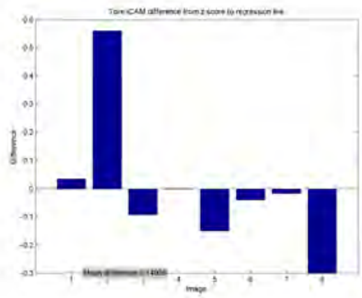


(d) Cartoon.

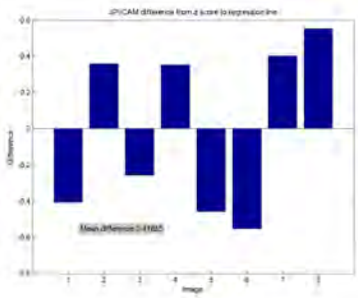
Figure 177: Difference from the regression line - S-CIELAB.



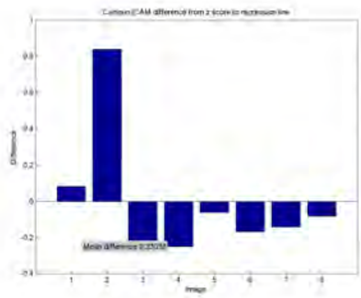
(a) Girl.



(b) Tore.

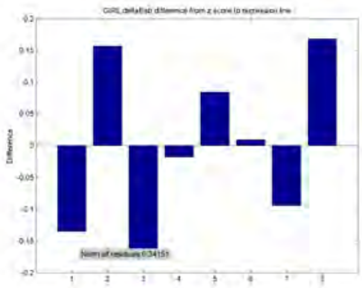


(c) JP.

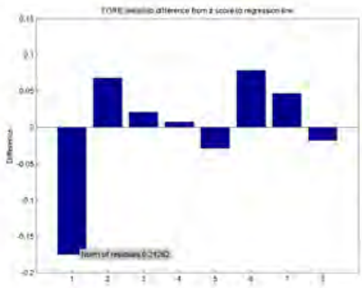


(d) Cartoon.

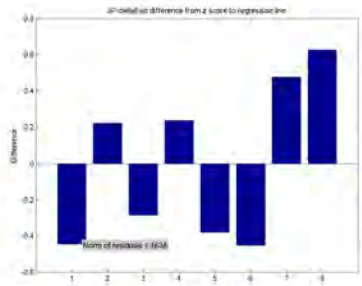
Figure 178: Difference from the regression line - iCAM.



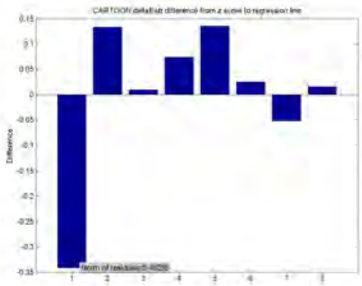
(a) Girl.



(b) Tore.

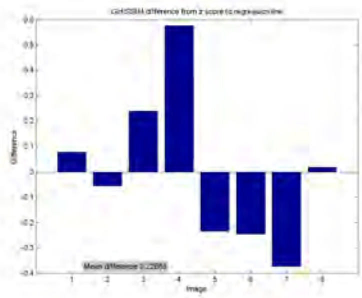


(c) JP.

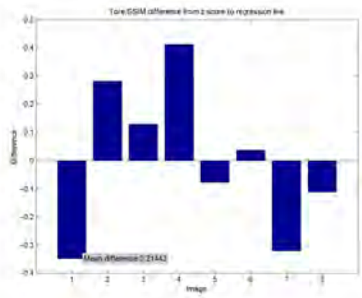


(d) Cartoon.

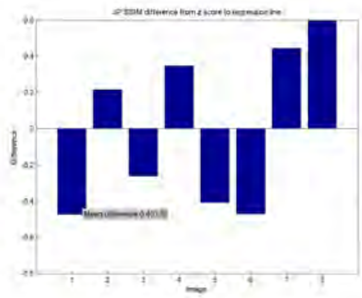
Figure 179: Difference from the regression line - ΔE_{ab}.



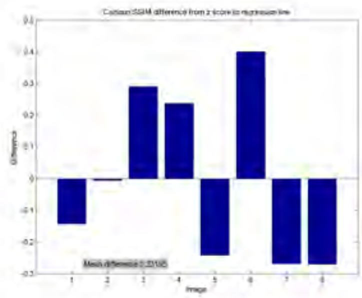
(a) Girl.



(b) Tore.

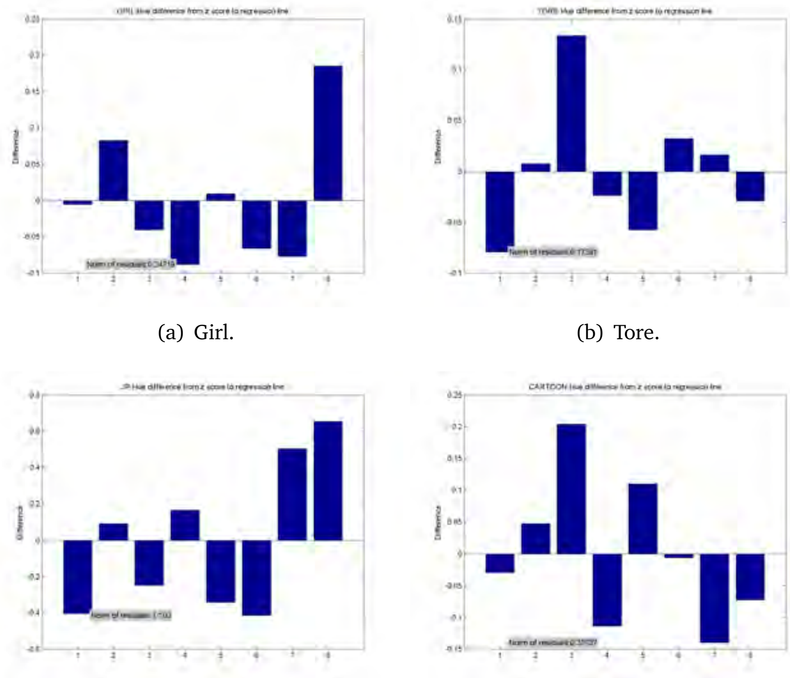


(c) JP.



(d) Cartoon.

Figure 180: Difference from the regression line - SSIM.



(a) Girl. (b) Tore.

(c) JP. (d) Cartoon.

Figure 181: Difference from the regression line - hue angle algorithm.

D.2 Image difference metrics applied to observer stated regions

Scene	Mean
CARTOON_LM3.PNG	0.0505
CARTOON_LM5.PNG	0.0815
CARTOON_LP3.PNG	0.0512
CARTOON_LP5.PNG	0.0828
CARTOON_REGION_SIGN_LM3.PNG	0.0060
CARTOON_REGION_SIGN_LP3.PNG	0.0062
CARTOON_REGION_TSHIRT_LM3.PNG	0.0029
CARTOON_REGION_TSHIRT_LP3.PNG	0.0034
GIRL_LM3.PNG	0.0342
GIRL_LM5.PNG	0.0552
GIRL_LP3.PNG	0.0345
GIRL_LP5.PNG	0.0558
GIRL_REGION_BACK_LM3.PNG	0.0181
GIRL_REGION_BACK_LP3.PNG	0.0182
GIRL_REGION_FRONT_LM3.PNG	0.0195
GIRL_REGION_FRONT_LP3.PNG	0.0200
JP_LM3.PNG	0.0248
JP_LM5.PNG	0.0425
JP_LP3.PNG	0.0252
JP_LP5.PNG	0.0406
JP_REGION_STRIPE2_LM3.PNG	0.0040
JP_REGION_STRIPE2_LP3.PNG	0.0041
JP_REGION_STRIPE_LM3.PNG	0.0066
JP_REGION_STRIPE_LP3.PNG	0.0069
TORE_LM3.PNG	0.0321
TORE_LM5.PNG	0.0530
TORE_LP3.PNG	0.0285
TORE_LP5.PNG	0.0456
TORE_REGION_BACK_LM3.PNG	0.0123
TORE_REGION_BACK_LP3.PNG	0.0111
TORE_REGION_FRONT_LM3.PNG	0.0231
TORE_REGION_FRONT_LP3.PNG	0.0204

Table 40: Observer regions: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.0054
CARTOON_LM5.PNG	w
CARTOON_LP3.PNG	0.0109
CARTOON_LP5.PNG	0.0179
CARTOON_REGION_SIGN_LM3.PNG	0.0017
CARTOON_REGION_SIGN_LP3.PNG	0.0017
CARTOON_REGION_TSHIRT_LM3.PNG	0.0005
CARTOON_REGION_TSHIRT_LP3.PNG	0.0005
GIRL_LM3.PNG	0.0046
GIRL_LM5.PNG	0.0070
GIRL_LP3.PNG	0.0080
GIRL_LP5.PNG	0.0128
GIRL_REGION_BACK_LM3.PNG	0.0041
GIRL_REGION_BACK_LP3.PNG	0.0042
GIRL_REGION_FRONT_LM3.PNG	0.0050
GIRL_REGION_FRONT_LP3.PNG	0.0063
JP_LM3.PNG	0.0032
JP_LM5.PNG	0.0047
JP_LP3.PNG	0.0032
JP_LP5.PNG	0.0044
JP_REGION_STRIPE2_LM3.PNG	0.0018
JP_REGION_STRIPE2_LP3.PNG	0.0030
JP_REGION_STRIPE_LM3.PNG	0.0019
JP_REGION_STRIPE_LP3.PNG	0.0019
TORE_LM3.PNG	0.0039
TORE_LM5.PNG	0.0059
TORE_LP3.PNG	0.0062
TORE_LP5.PNG	0.0098
TORE_REGION_BACK_LM3.PNG	0.0033
TORE_REGION_BACK_LP3.PNG	0.0028
TORE_REGION_FRONT_LM3.PNG	0.0049
TORE_REGION_FRONT_LP3.PNG	0.0053

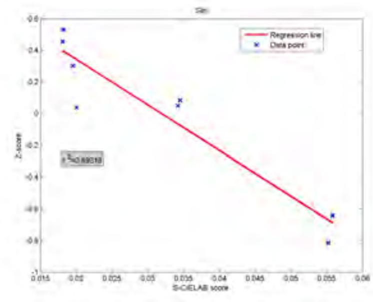
Table 41: Observer regions: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.0525
CARTOON_LM5.PNG	0.0816
CARTOON_LP3.PNG	0.0458
CARTOON_LP5.PNG	0.0767
CARTOON_REGION_SIGN_LM3.PNG	0.0031
CARTOON_REGION_SIGN_LP3.PNG	0.0029
CARTOON_REGION_TSHIRT_LM3.PNG	0.0007
CARTOON_REGION_TSHIRT_LP3.PNG	0.0007
GIRL_LM3.PNG	0.0304
GIRL_LM5.PNG	0.0494
GIRL_LP3.PNG	0.0304
GIRL_LP5.PNG	0.0494
GIRL_REGION_BACK_LM3.PNG	0.0158
GIRL_REGION_BACK_LP3.PNG	0.0158
GIRL_REGION_FRONT_LM3.PNG	0.0146
GIRL_REGION_FRONT_LP3.PNG	0.0146
JP_LM3.PNG	0.0211
JP_LM5.PNG	0.0367
JP_LP3.PNG	0.0206
JP_LP5.PNG	0.0334
JP_REGION_STRIPE2_LM3.PNG	0.0023
JP_REGION_STRIPE2_LP3.PNG	0.0023
JP_REGION_STRIPE_LM3.PNG	0.0033
JP_REGION_STRIPE_LP3.PNG	0.0034
TORE_LM3.PNG	0.0275
TORE_LM5.PNG	0.0458
TORE_LP3.PNG	0.0244
TORE_LP5.PNG	0.0394
TORE_REGION_BACK_LM3.PNG	0.0102
TORE_REGION_BACK_LP3.PNG	0.0090
TORE_REGION_FRONT_LM3.PNG	0.0173
TORE_REGION_FRONT_LP3.PNG	0.0154

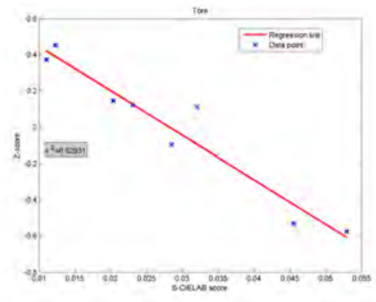
Table 42: Observer regions: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.01390
CARTOON_LM5.PNG	0.01384
CARTOON_LP3.PNG	0.01390
CARTOON_LP5.PNG	0.01386
CARTOON_REGION_SIGN_LM3.PNG	0.01386
CARTOON_REGION_SIGN_LP3.PNG	0.01393
CARTOON_REGION_TSHIRT_LM3.PNG	0.01393
CARTOON_REGION_TSHIRT_LP3.PNG	0.01392
GIRL_LM3.PNG	0.00964
GIRL_LM5.PNG	0.00957
GIRL_LP3.PNG	0.00965
GIRL_LP5.PNG	0.00962
GIRL_REGION_BACK_LM3.PNG	0.00963
GIRL_REGION_BACK_LP3.PNG	0.00963
GIRL_REGION_FRONT_LM3.PNG	0.00962
GIRL_REGION_FRONT_LP3.PNG	0.00962
JP_LM3.PNG	0.00623
JP_LM5.PNG	0.00598
JP_LP3.PNG	0.00631
JP_LP5.PNG	0.00626
JP_REGION_STRIPE2_LM3.PNG	0.00661
JP_REGION_STRIPE2_LP3.PNG	0.00665
JP_REGION_STRIPE_LM3.PNG	0.00655
JP_REGION_STRIPE_LP3.PNG	0.00655
TORE_LM3.PNG	0.00784
TORE_LM5.PNG	0.00777
TORE_LP3.PNG	0.00783
TORE_LP5.PNG	0.00768
TORE_REGION_BACK_LM3.PNG	0.00789
TORE_REGION_BACK_LP3.PNG	0.00790
TORE_REGION_FRONT_LM3.PNG	0.00785
TORE_REGION_FRONT_LP3.PNG	0.00783

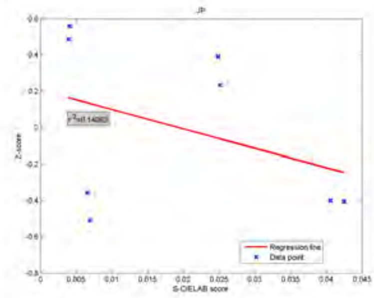
Table 43: Observer regions: SSIM score for all scenes and reproductions.



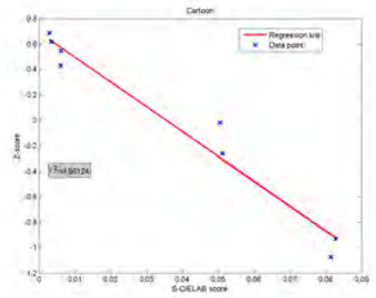
(a) Girl.



(b) Tore.

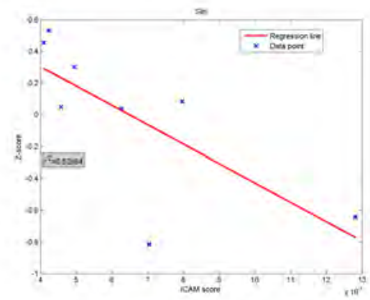


(c) JP.

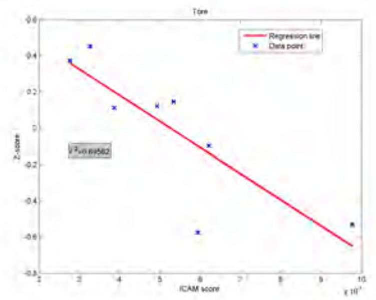


(d) Cartoon.

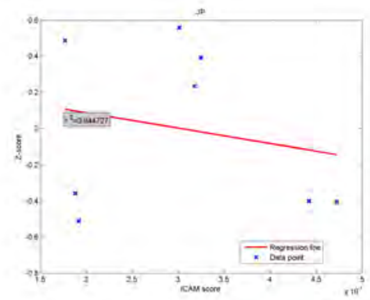
Figure 182: Observer regions: S-CIELAB regression plot for each scene.



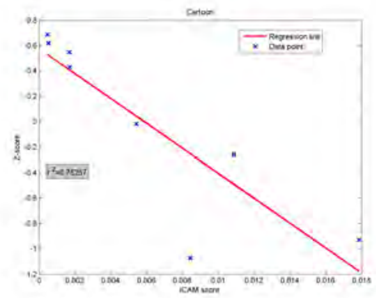
(a) Girl.



(b) Tore.

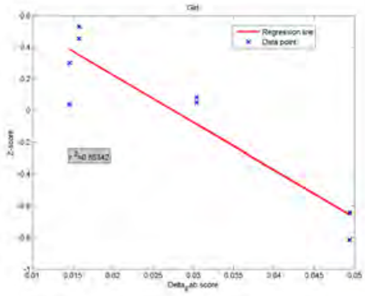


(c) JP.

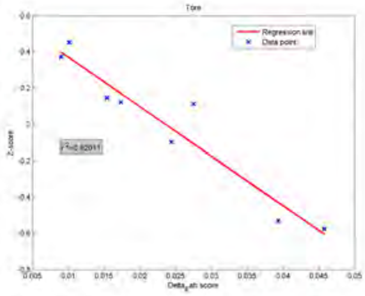


(d) Cartoon.

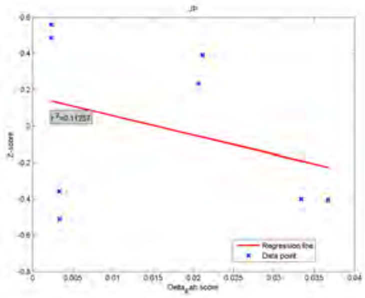
Figure 183: Observer regions: iCAM regression plot for each scene.



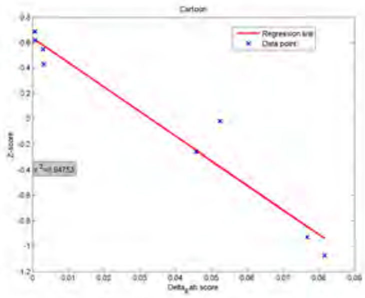
(a) Girl.



(b) Tore.

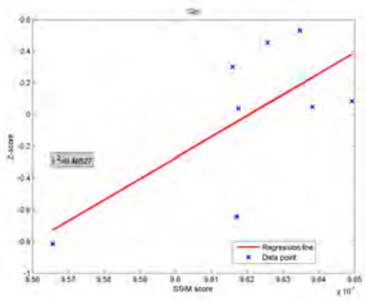


(c) JP.

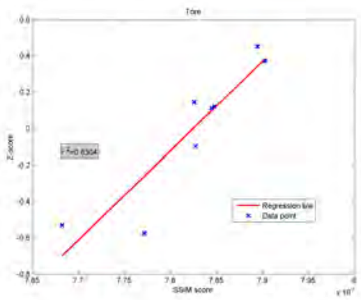


(d) Cartoon.

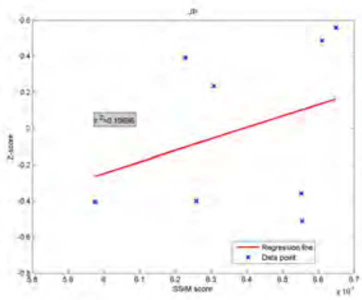
Figure 184: Observer regions: ΔE_{ab} regression plot for each scene.



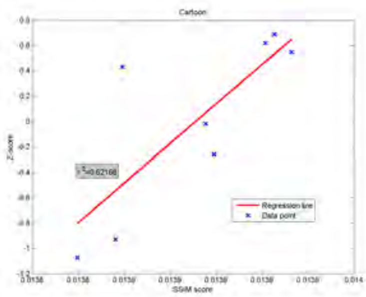
(a) Girl.



(b) Tore.

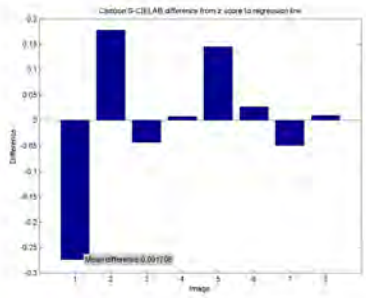


(c) JP.

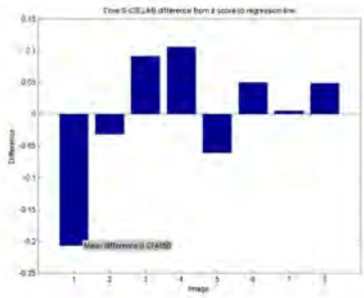


(d) Cartoon.

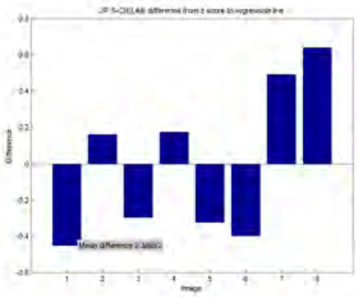
Figure 185: Observer regions: SSIM regression plot for each scene.



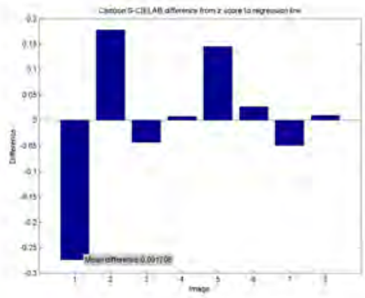
(a) Girl.



(b) Tore.

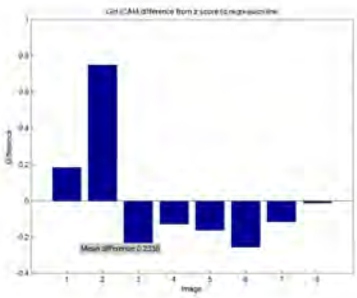


(c) JP.

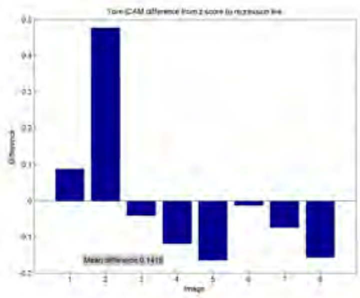


(d) Cartoon.

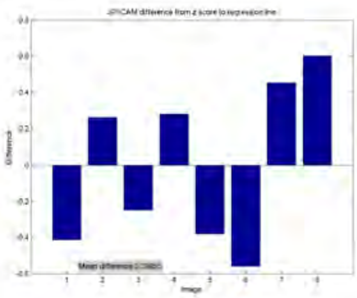
Figure 186: Observer regions: Difference from the regression line - S-CIELAB.



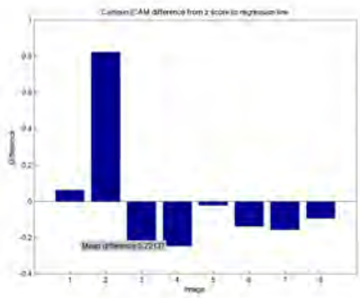
(a) Girl.



(b) Tore.

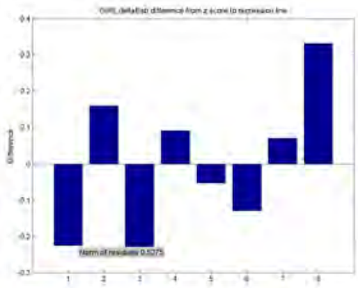


(c) JP.

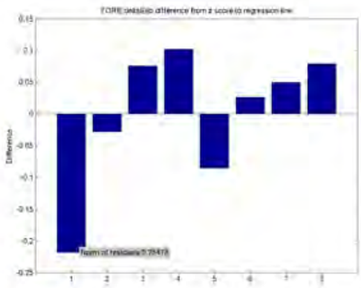


(d) Cartoon.

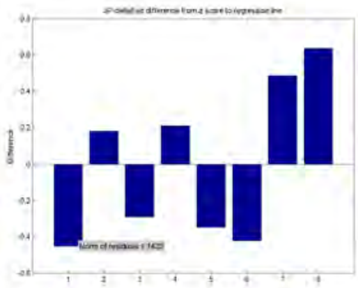
Figure 187: Observer regions: Difference from the regression line - iCAM.



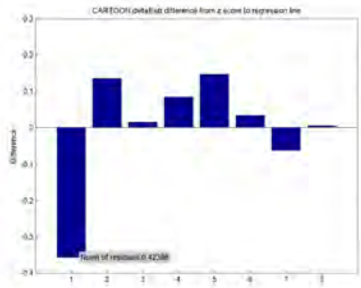
(a) Girl.



(b) Tore.

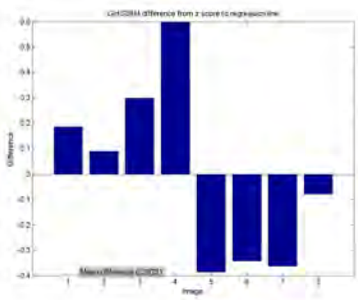


(c) JP.

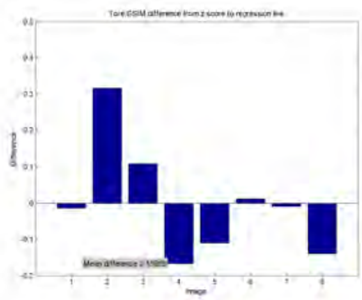


(d) Cartoon.

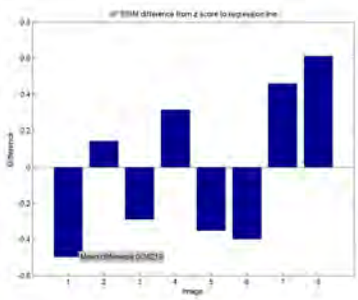
Figure 188: Observer regions: Difference from the regression line - ΔE_{ab}.



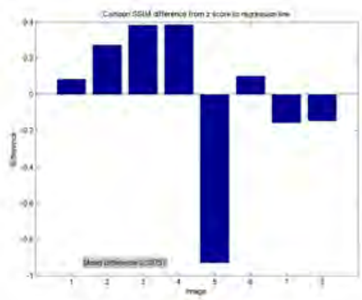
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 189: Observer regions: Difference from the regression line - SSIM.

D.3 Image difference metrics applied to eye tracker regions

Scene	Mean
CARTOON_LM3.PNG	0.7653
CARTOON_LM5.PNG	1.2147
CARTOON_LP3.PNG	0.7928
CARTOON_LP5.PNG	1.2773
CARTOON_REGION_SIGN_LM3.PNG	0.1626
CARTOON_REGION_SIGN_LP3.PNG	0.1660
CARTOON_REGION_TSHIRT_LM3.PNG	0.1346
CARTOON_REGION_TSHIRT_LP3.PNG	0.1566
GIRL_LM3.PNG	0.6103
GIRL_LM5.PNG	0.9735
GIRL_LP3.PNG	0.6225
GIRL_LP5.PNG	1.0045
GIRL_REGION_BACK_LM3.PNG	0.2064
GIRL_REGION_BACK_LP3.PNG	0.2102
GIRL_REGION_FRONT_LM3.PNG	0.5163
GIRL_REGION_FRONT_LP3.PNG	0.5326
JP_LM3.PNG	0.5859
JP_LM5.PNG	0.9971
JP_LP3.PNG	0.6071
JP_LP5.PNG	0.9778
JP_REGION_STRIPE2_LM3.PNG	0.0425
JP_REGION_STRIPE2_LP3.PNG	0.0439
JP_REGION_STRIPE_LM3.PNG	0.1854
JP_REGION_STRIPE_LP3.PNG	0.1953
TORE_LM3.PNG	0.4012
TORE_LM5.PNG	0.6585
TORE_LP3.PNG	0.3600
TORE_LP5.PNG	0.5760
TORE_REGION_BACK_LM3.PNG	0.0886
TORE_REGION_BACK_LP3.PNG	0.0817
TORE_REGION_FRONT_LM3.PNG	0.3618
TORE_REGION_FRONT_LP3.PNG	0.3211

Table 44: Eye tracker regions: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.0825
CARTOON_LM5.PNG	0.1249
CARTOON_LP3.PNG	0.1623
CARTOON_LP5.PNG	0.2633
CARTOON_REGION_SIGN_LM3.PNG	0.0424
CARTOON_REGION_SIGN_LP3.PNG	0.0413
CARTOON_REGION_TSHIRT_LM3.PNG	0.0230
CARTOON_REGION_TSHIRT_LP3.PNG	0.0246
GIRL_LM3.PNG	0.1102
GIRL_LM5.PNG	0.1679
GIRL_LP3.PNG	0.1543
GIRL_LP5.PNG	0.2412
GIRL_REGION_BACK_LM3.PNG	0.0752
GIRL_REGION_BACK_LP3.PNG	0.0774
GIRL_REGION_FRONT_LM3.PNG	0.1417
GIRL_REGION_FRONT_LP3.PNG	0.1677
JP_LM3.PNG	0.0759
JP_LM5.PNG	0.1110
JP_LP3.PNG	0.0901
JP_LP5.PNG	0.1210
JP_REGION_STRIPE2_LM3.PNG	0.0242
JP_REGION_STRIPE2_LP3.PNG	0.0574
JP_REGION_STRIPE_LM3.PNG	0.0498
JP_REGION_STRIPE_LP3.PNG	0.0504
TORE_LM3.PNG	0.0619
TORE_LM5.PNG	0.0925
TORE_LP3.PNG	0.0884
TORE_LP5.PNG	0.1362
TORE_REGION_BACK_LM3.PNG	0.0284
TORE_REGION_BACK_LP3.PNG	0.0252
TORE_REGION_FRONT_LM3.PNG	0.0699
TORE_REGION_FRONT_LP3.PNG	0.0868

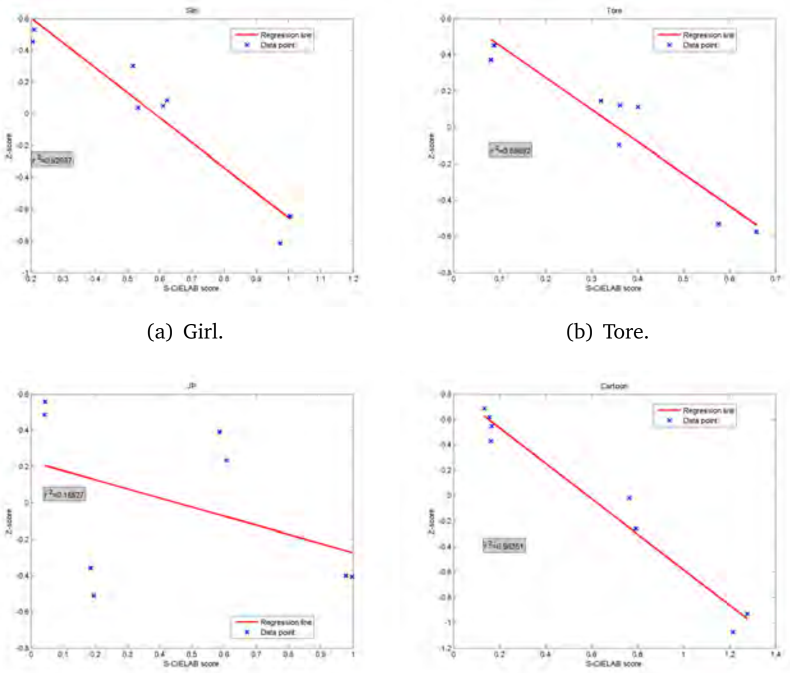
Table 45: Eye tracker regions: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.7076
CARTOON_LM5.PNG	1.1137
CARTOON_LP3.PNG	0.6380
CARTOON_LP5.PNG	1.0594
CARTOON_REGION_SIGN_LM3.PNG	0.0856
CARTOON_REGION_SIGN_LP3.PNG	0.0807
CARTOON_REGION_TSHIRT_LM3.PNG	0.0345
CARTOON_REGION_TSHIRT_LP3.PNG	0.0349
GIRL_LM3.PNG	0.4972
GIRL_LM5.PNG	0.8067
GIRL_LP3.PNG	0.4972
GIRL_LP5.PNG	0.8080
GIRL_REGION_BACK_LM3.PNG	0.1261
GIRL_REGION_BACK_LP3.PNG	0.1261
GIRL_REGION_FRONT_LM3.PNG	0.3711
GIRL_REGION_FRONT_LP3.PNG	0.3712
JP_LM3.PNG	0.4691
JP_LM5.PNG	0.8135
JP_LP3.PNG	0.4748
JP_LP5.PNG	0.7692
JP_REGION_STRIPE2_LM3.PNG	0.0239
JP_REGION_STRIPE2_LP3.PNG	0.0243
JP_REGION_STRIPE_LM3.PNG	0.0871
JP_REGION_STRIPE_LP3.PNG	0.0880
TORE_LM3.PNG	0.3221
TORE_LM5.PNG	0.5351
TORE_LP3.PNG	0.2897
TORE_LP5.PNG	0.4689
TORE_REGION_BACK_LM3.PNG	0.0536
TORE_REGION_BACK_LP3.PNG	0.0473
TORE_REGION_FRONT_LM3.PNG	0.2685
TORE_REGION_FRONT_LP3.PNG	0.2424

Table 46: Eye tracker regions: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.20394
CARTOON_LM5.PNG	0.20236
CARTOON_LP3.PNG	0.20435
CARTOON_LP5.PNG	0.20316
CARTOON_REGION_SIGN_LM3.PNG	0.20591
CARTOON_REGION_SIGN_LP3.PNG	0.20627
CARTOON_REGION_TSHIRT_LM3.PNG	0.20473
CARTOON_REGION_TSHIRT_LP3.PNG	0.20477
GIRL_LM3.PNG	0.16295
GIRL_LM5.PNG	0.16012
GIRL_LP3.PNG	0.16351
GIRL_LP5.PNG	0.16266
GIRL_REGION_BACK_LM3.PNG	0.16373
GIRL_REGION_BACK_LP3.PNG	0.16381
GIRL_REGION_FRONT_LM3.PNG	0.16275
GIRL_REGION_FRONT_LP3.PNG	0.16323
JP_LM3.PNG	0.14436
JP_LM5.PNG	0.13734
JP_LP3.PNG	0.14618
JP_LP5.PNG	0.14465
JP_REGION_STRIPE2_LM3.PNG	0.15545
JP_REGION_STRIPE2_LP3.PNG	0.15588
JP_REGION_STRIPE_LM3.PNG	0.15350
JP_REGION_STRIPE_LP3.PNG	0.15369
TORE_LM3.PNG	0.09390
TORE_LM5.PNG	0.09286
TORE_LP3.PNG	0.09487
TORE_LP5.PNG	0.09335
TORE_REGION_BACK_LM3.PNG	0.09568
TORE_REGION_BACK_LP3.PNG	0.09596
TORE_REGION_FRONT_LM3.PNG	0.09411
TORE_REGION_FRONT_LP3.PNG	0.09484

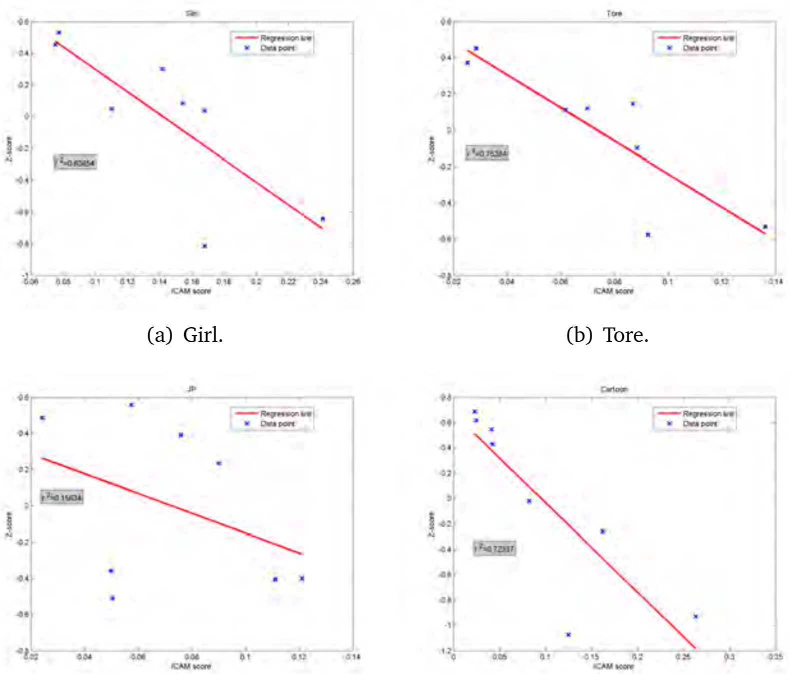
Table 47: Eye tracker regions: SSIM score for all scenes and reproductions.



(a) Girl. (b) Tore.

(c) JP. (d) Cartoon.

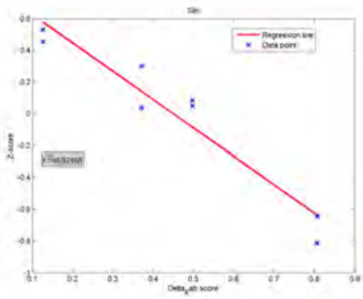
Figure 190: Eye tracker regions: S-CIELAB regression plot for each scene.



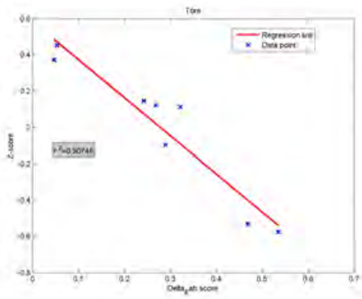
(a) Girl. (b) Tore.

(c) JP. (d) Cartoon.

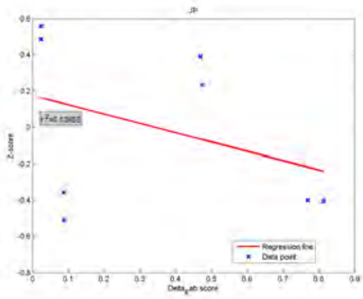
Figure 191: Eye tracker regions: iCAM regression plot for each scene.



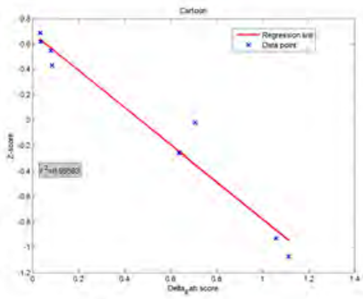
(a) Girl.



(b) Tore.

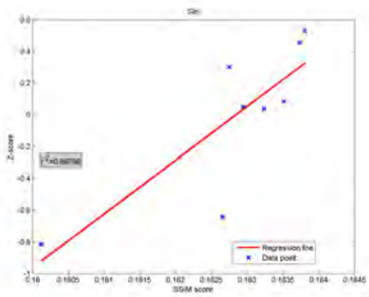


(c) JP.

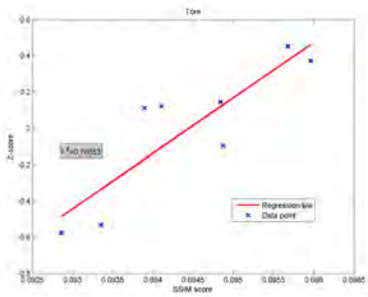


(d) Cartoon.

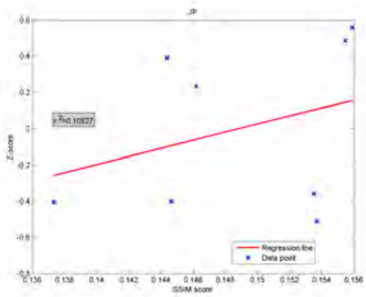
Figure 192: Eye tracker regions: ΔE_{ab} regression plot for each scene.



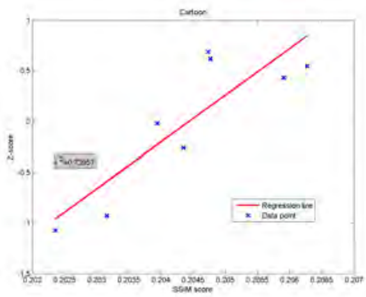
(a) Girl.



(b) Tore.

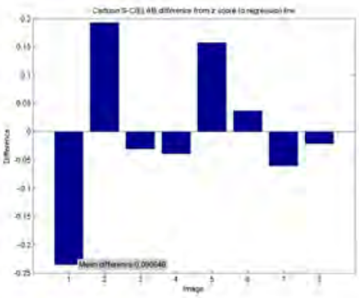


(c) JP.

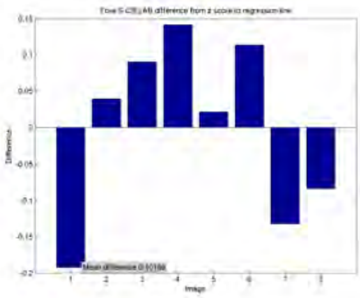


(d) Cartoon.

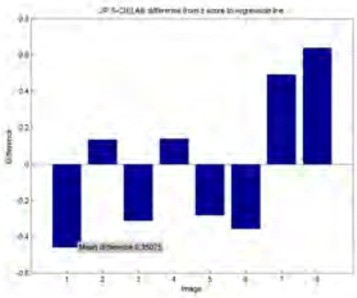
Figure 193: Eye tracker regions: SSIM regression plot for each scene.



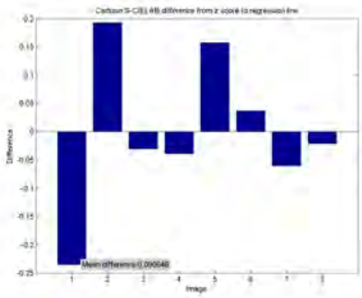
(a) Girl.



(b) Tore.

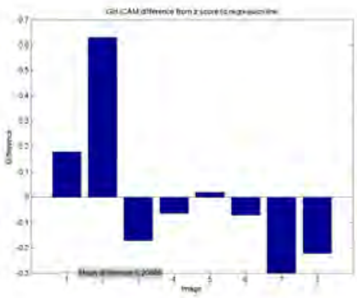


(c) JP.

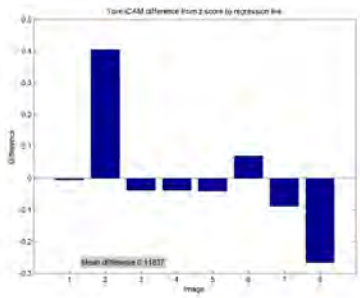


(d) Cartoon.

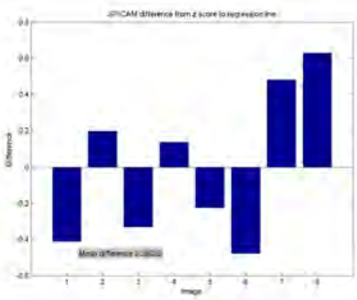
Figure 194: Eye tracker regions: Difference from the regression line - S-CIELAB.



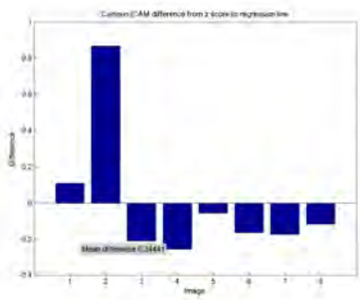
(a) Girl.



(b) Tore.

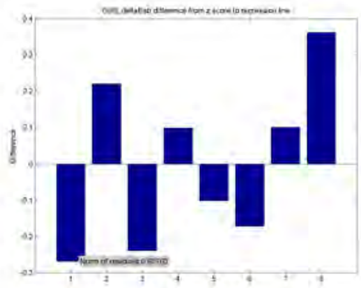


(c) JP.

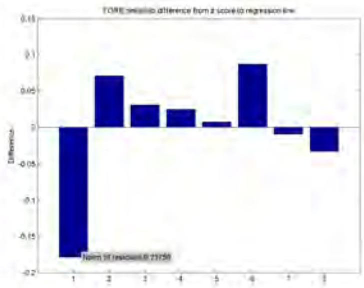


(d) Cartoon.

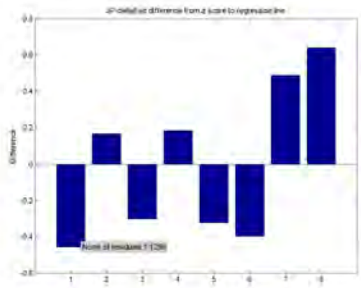
Figure 195: Eye tracker regions: Difference from the regression line - iCAM.



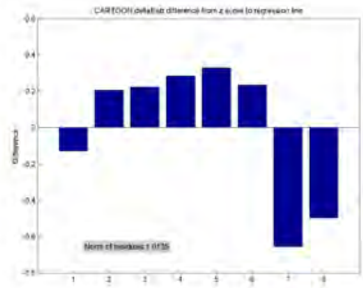
(a) Girl.



(b) Tore.

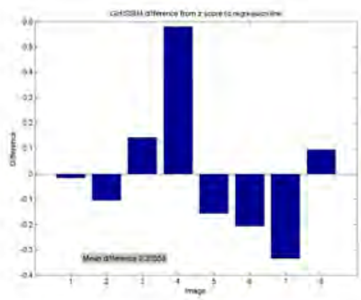


(c) JP.

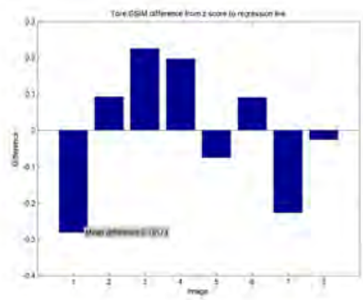


(d) Cartoon.

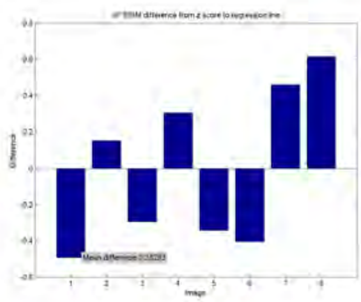
Figure 196: Eye tracker regions: Difference from the regression line - ΔE_{ab}.



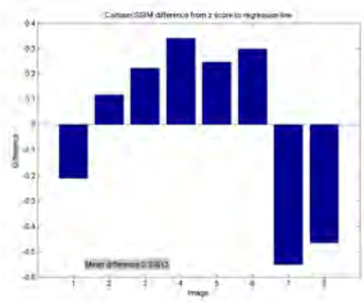
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 197: Eye tracker regions: Difference from the regression line - SSIM.

D.4 Image difference metrics applied to eye tracker regions part 2

Scene	Mean
CARTOON_LM3.PNG	0.353
CARTOON_LM5.PNG	0.562
CARTOON_LP3.PNG	0.364
CARTOON_LP5.PNG	0.587
CARTOON_REGION_SIGN_LM3.PNG	0.079
CARTOON_REGION_SIGN_LP3.PNG	0.079
CARTOON_REGION_TSHIRT_LM3.PNG	0.054
CARTOON_REGION_TSHIRT_LP3.PNG	0.064
GIRL_LM3.PNG	0.397
GIRL_LM5.PNG	0.634
GIRL_LP3.PNG	0.405
GIRL_LP5.PNG	0.653
GIRL_REGION_BACK_LM3.PNG	0.144
GIRL_REGION_BACK_LP3.PNG	0.146
GIRL_REGION_FRONT_LM3.PNG	0.328
GIRL_REGION_FRONT_LP3.PNG	0.338
JP_LM3.PNG	0.296
JP_LM5.PNG	0.503
JP_LP3.PNG	0.304
JP_LP5.PNG	0.489
JP_REGION_STRIPE2_LM3.PNG	0.035
JP_REGION_STRIPE2_LP3.PNG	0.037
JP_REGION_STRIPE_LM3.PNG	0.082
JP_REGION_STRIPE_LP3.PNG	0.086
TORE_LM3.PNG	0.285
TORE_LM5.PNG	0.467
TORE_LP3.PNG	0.257
TORE_LP5.PNG	0.411
TORE_REGION_BACK_LM3.PNG	0.082
TORE_REGION_BACK_LP3.PNG	0.076
TORE_REGION_FRONT_LM3.PNG	0.241
TORE_REGION_FRONT_LP3.PNG	0.215

Table 48: Eye tracker regions part 2: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.038
CARTOON_LM5.PNG	0.058
CARTOON_LP3.PNG	0.075
CARTOON_LP5.PNG	0.121
CARTOON_REGION_SIGN_LM3.PNG	0.020
CARTOON_REGION_SIGN_LP3.PNG	0.020
CARTOON_REGION_TSHIRT_LM3.PNG	0.009
CARTOON_REGION_TSHIRT_LP3.PNG	0.010
GIRL_LM3.PNG	0.070
GIRL_LM5.PNG	0.107
GIRL_LP3.PNG	0.100
GIRL_LP5.PNG	0.157
GIRL_REGION_BACK_LM3.PNG	0.048
GIRL_REGION_BACK_LP3.PNG	0.050
GIRL_REGION_FRONT_LM3.PNG	0.089
GIRL_REGION_FRONT_LP3.PNG	0.105
JP_LM3.PNG	0.041
JP_LM5.PNG	0.059
JP_LP3.PNG	0.047
JP_LP5.PNG	0.063
JP_REGION_STRIPE2_LM3.PNG	0.018
JP_REGION_STRIPE2_LP3.PNG	0.034
JP_REGION_STRIPE_LM3.PNG	0.023
JP_REGION_STRIPE_LP3.PNG	0.024
TORE_LM3.PNG	0.046
TORE_LM5.PNG	0.068
TORE_LP3.PNG	0.069
TORE_LP5.PNG	0.105
TORE_REGION_BACK_LM3.PNG	0.027
TORE_REGION_BACK_LP3.PNG	0.024
TORE_REGION_FRONT_LM3.PNG	0.053
TORE_REGION_FRONT_LP3.PNG	0.065

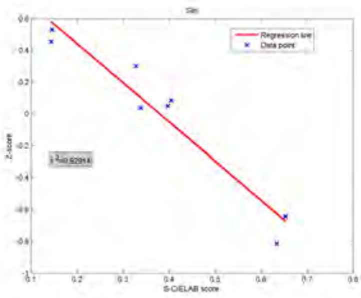
Table 49: Eye tracker regions part 2: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.332
CARTOON_LM5.PNG	0.520
CARTOON_LP3.PNG	0.297
CARTOON_LP5.PNG	0.494
CARTOON_REGION_SIGN_LM3.PNG	0.040
CARTOON_REGION_SIGN_LP3.PNG	0.037
CARTOON_REGION_TSHIRT_LM3.PNG	0.014
CARTOON_REGION_TSHIRT_LP3.PNG	0.014
GIRL_LM3.PNG	0.325
GIRL_LM5.PNG	0.527
GIRL_LP3.PNG	0.325
GIRL_LP5.PNG	0.527
GIRL_REGION_BACK_LM3.PNG	0.092
GIRL_REGION_BACK_LP3.PNG	0.092
GIRL_REGION_FRONT_LM3.PNG	0.232
GIRL_REGION_FRONT_LP3.PNG	0.233
JP_LM3.PNG	0.239
JP_LM5.PNG	0.413
JP_LP3.PNG	0.241
JP_LP5.PNG	0.390
JP_REGION_STRIPE2_LM3.PNG	0.022
JP_REGION_STRIPE2_LP3.PNG	0.023
JP_REGION_STRIPE_LM3.PNG	0.039
JP_REGION_STRIPE_LP3.PNG	0.039
TORE_LM3.PNG	0.234
TORE_LM5.PNG	0.388
TORE_LP3.PNG	0.214
TORE_LP5.PNG	0.345
TORE_REGION_BACK_LM3.PNG	0.056
TORE_REGION_BACK_LP3.PNG	0.050
TORE_REGION_FRONT_LM3.PNG	0.178
TORE_REGION_FRONT_LP3.PNG	0.164

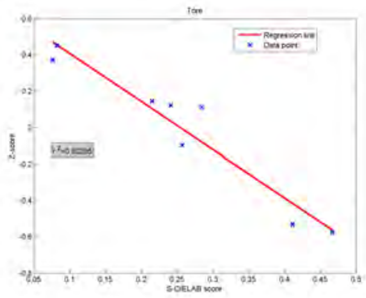
Table 50: Eye tracker regions part 2: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.093
CARTOON_LM5.PNG	0.092
CARTOON_LP3.PNG	0.093
CARTOON_LP5.PNG	0.093
CARTOON_REGION_SIGN_LM3.PNG	0.094
CARTOON_REGION_SIGN_LP3.PNG	0.094
CARTOON_REGION_TSHIRT_LM3.PNG	0.093
CARTOON_REGION_TSHIRT_LP3.PNG	0.093
GIRL_LM3.PNG	0.106
GIRL_LM5.PNG	0.104
GIRL_LP3.PNG	0.106
GIRL_LP5.PNG	0.105
GIRL_REGION_BACK_LM3.PNG	0.106
GIRL_REGION_BACK_LP3.PNG	0.106
GIRL_REGION_FRONT_LM3.PNG	0.106
GIRL_REGION_FRONT_LP3.PNG	0.106
JP_LM3.PNG	0.072
JP_LM5.PNG	0.069
JP_LP3.PNG	0.074
JP_LP5.PNG	0.073
JP_REGION_STRIPE2_LM3.PNG	0.078
JP_REGION_STRIPE2_LP3.PNG	0.078
JP_REGION_STRIPE_LM3.PNG	0.077
JP_REGION_STRIPE_LP3.PNG	0.077
TORE_LM3.PNG	0.067
TORE_LM5.PNG	0.067
TORE_LP3.PNG	0.069
TORE_LP5.PNG	0.067
TORE_REGION_BACK_LM3.PNG	0.070
TORE_REGION_BACK_LP3.PNG	0.070
TORE_REGION_FRONT_LM3.PNG	0.068
TORE_REGION_FRONT_LP3.PNG	0.069

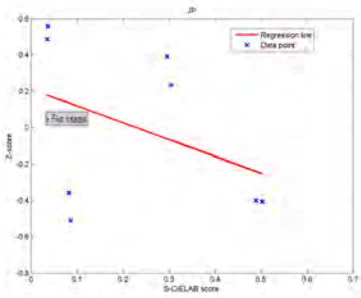
Table 51: Eye tracker regions part 2: SSIM score for all scenes and reproductions.



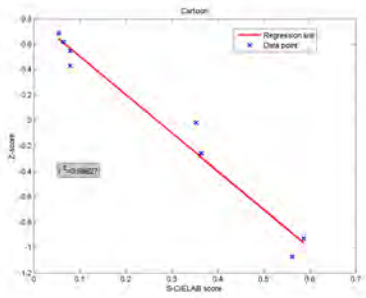
(a) Girl.



(b) Tore.

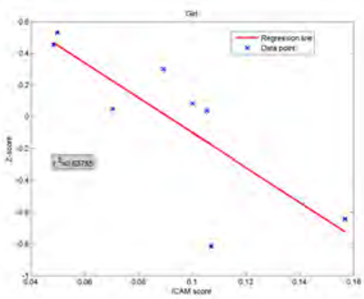


(c) JP.

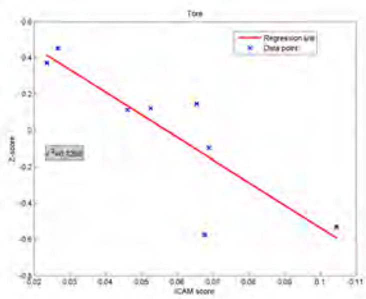


(d) Cartoon.

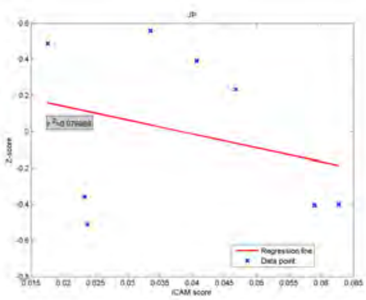
Figure 198: Eye tracker regions part 2: S-CIELAB regression plot for each scene.



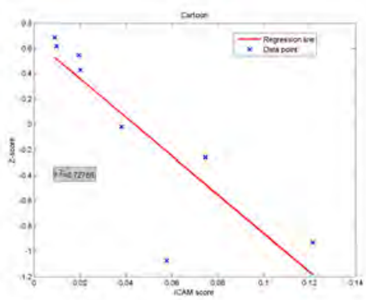
(a) Girl.



(b) Tore.

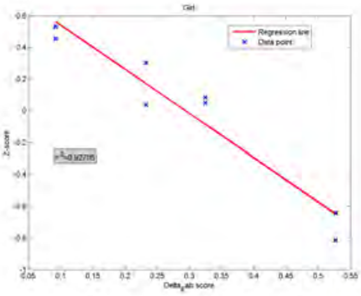


(c) JP.

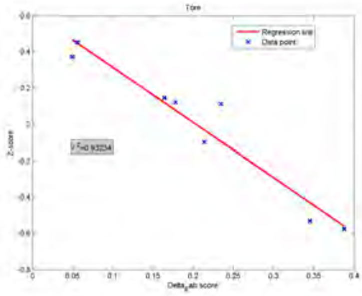


(d) Cartoon.

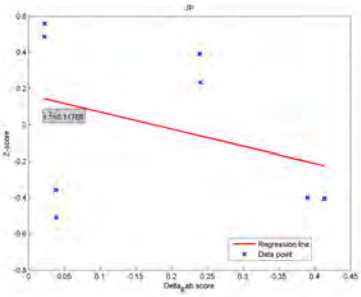
Figure 199: Eye tracker regions part 2: iCAM regression plot for each scene.



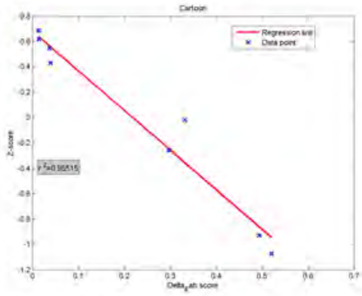
(a) Girl.



(b) Tore.

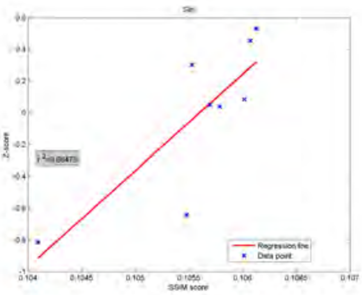


(c) JP.

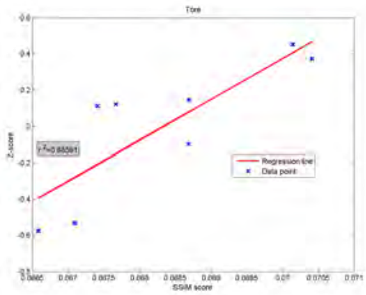


(d) Cartoon.

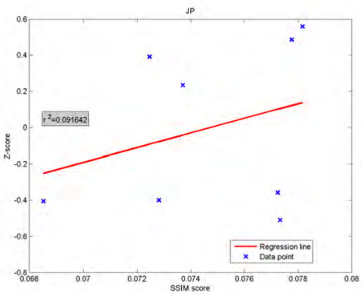
Figure 200: Eye tracker regions part 2: ΔE_{ab} regression plot for each scene.



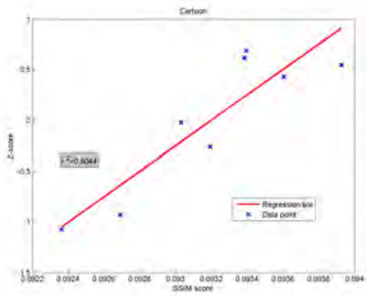
(a) Girl.



(b) Tore.

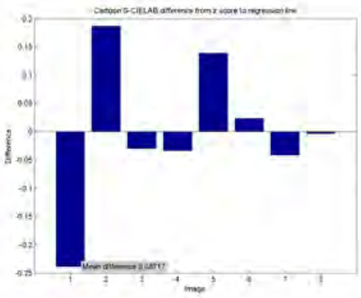


(c) JP.

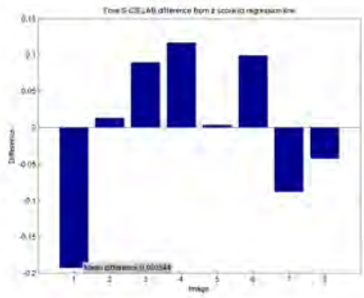


(d) Cartoon.

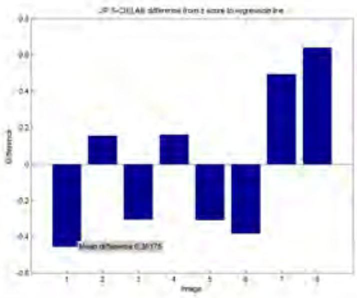
Figure 201: Eye tracker regions part 2: SSIM regression plot for each scene.



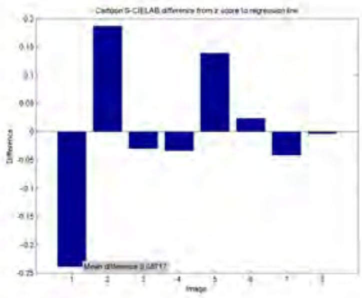
(a) Girl.



(b) Tore.

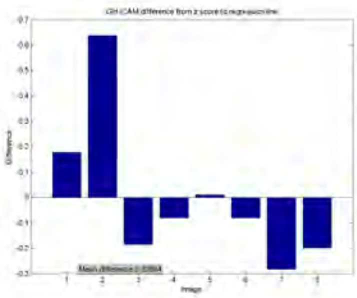


(c) JP.

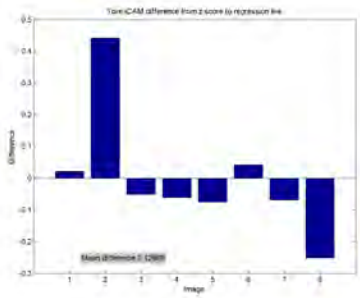


(d) Cartoon.

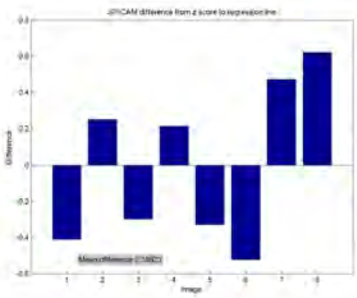
Figure 202: Eye tracker regions part 2: Difference from the regression line - S-CIELAB.



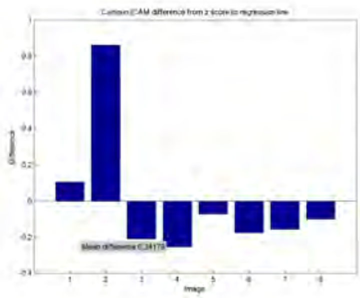
(a) Girl.



(b) Tore.

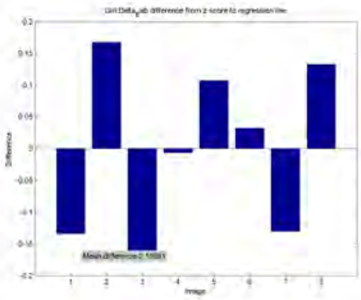


(c) JP.

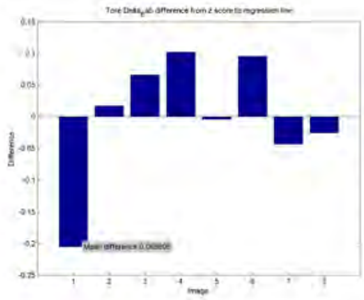


(d) Cartoon.

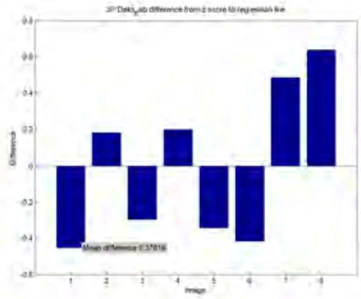
Figure 203: Eye tracker regions part 2: Difference from the regression line - iCAM.



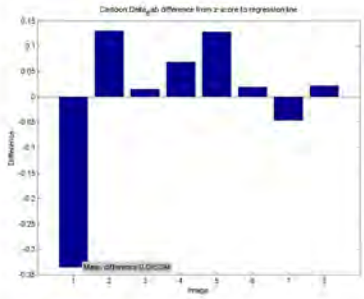
(a) Girl.



(b) Tore.

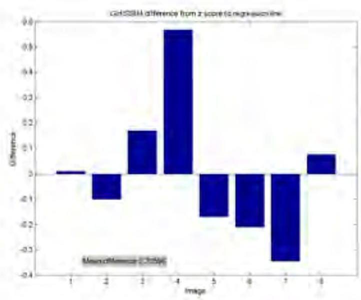


(c) JP.

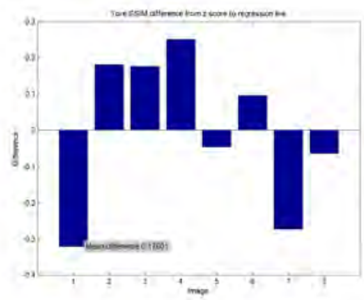


(d) Cartoon.

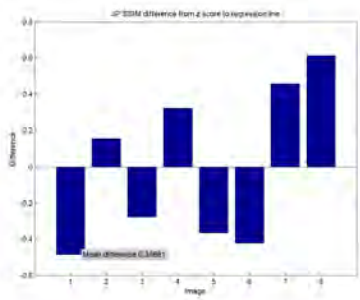
Figure 204: Eye tracker regions part 2: Difference from the regression line - ΔE_{ab}.



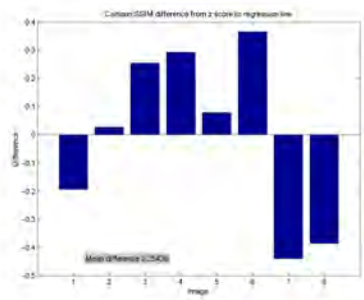
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 205: Eye tracker regions part 2: Difference from the regression line - SSIM.

D.5 Image difference metrics applied to combined eye tracker and observer regions

Scene	Mean
CARTOON_LM3.PNG	2.0285
CARTOON_LM5.PNG	3.2514
CARTOON_LP3.PNG	2.0720
CARTOON_LP5.PNG	3.3464
CARTOON_REGION_SIGN_LM3.PNG	0.3137
CARTOON_REGION_SIGN_LP3.PNG	0.3202
CARTOON_REGION_TSHIRT_LM3.PNG	0.2073
CARTOON_REGION_TSHIRT_LP3.PNG	0.2412
GIRL_LM3.PNG	1.4659
GIRL_LM5.PNG	2.3527
GIRL_LP3.PNG	1.4848
GIRL_LP5.PNG	2.4000
GIRL_REGION_BACK_LM3.PNG	0.6594
GIRL_REGION_BACK_LP3.PNG	0.6648
GIRL_REGION_FRONT_LM3.PNG	1.0047
GIRL_REGION_FRONT_LP3.PNG	1.0336
JP_LM3.PNG	1.2071
JP_LM5.PNG	2.0588
JP_LP3.PNG	1.2364
JP_LP5.PNG	1.9927
JP_REGION_STRIPE2_LM3.PNG	0.1413
JP_REGION_STRIPE2_LP3.PNG	0.1461
JP_REGION_STRIPE_LM3.PNG	0.3505
JP_REGION_STRIPE_LP3.PNG	0.3688
TORE_LM3.PNG	1.2033
TORE_LM5.PNG	1.9827
TORE_LP3.PNG	1.0722
TORE_LP5.PNG	1.7154
TORE_REGION_BACK_LM3.PNG	0.3961
TORE_REGION_BACK_LP3.PNG	0.3582
TORE_REGION_FRONT_LM3.PNG	0.9386
TORE_REGION_FRONT_LP3.PNG	0.8299

Table 52: Combined eye tracker and observer regions: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.2182
CARTOON_LM5.PNG	0.3356
CARTOON_LP3.PNG	0.4340
CARTOON_LP5.PNG	0.7096
CARTOON_REGION_SIGN_LM3.PNG	0.0851
CARTOON_REGION_SIGN_LP3.PNG	0.0837
CARTOON_REGION_TSHIRT_LM3.PNG	0.0348
CARTOON_REGION_TSHIRT_LP3.PNG	0.0378
GIRL_LM3.PNG	0.2246
GIRL_LM5.PNG	0.3439
GIRL_LP3.PNG	0.3537
GIRL_LP5.PNG	0.5615
GIRL_REGION_BACK_LM3.PNG	0.1775
GIRL_REGION_BACK_LP3.PNG	0.1834
GIRL_REGION_FRONT_LM3.PNG	0.2656
GIRL_REGION_FRONT_LP3.PNG	0.3245
JP_LM3.PNG	0.1570
JP_LM5.PNG	0.2291
JP_LP3.PNG	0.1696
JP_LP5.PNG	0.2316
JP_REGION_STRIPE2_LM3.PNG	0.0684
JP_REGION_STRIPE2_LP3.PNG	0.1327
JP_REGION_STRIPE_LM3.PNG	0.0969
JP_REGION_STRIPE_LP3.PNG	0.0983
TORE_LM3.PNG	0.1586
TORE_LM5.PNG	0.2410
TORE_LP3.PNG	0.2438
TORE_LP5.PNG	0.3804
TORE_REGION_BACK_LM3.PNG	0.1101
TORE_REGION_BACK_LP3.PNG	0.0945
TORE_REGION_FRONT_LM3.PNG	0.1931
TORE_REGION_FRONT_LP3.PNG	0.2203

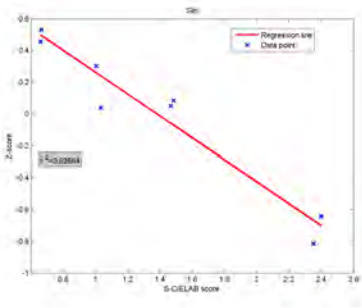
Table 53: Combined eye tracker and observer regions: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	2.0189
CARTOON_LM5.PNG	3.1535
CARTOON_LP3.PNG	1.7822
CARTOON_LP5.PNG	2.9779
CARTOON_REGION_SIGN_LM3.PNG	0.1634
CARTOON_REGION_SIGN_LP3.PNG	0.1543
CARTOON_REGION_TSHIRT_LM3.PNG	0.0524
CARTOON_REGION_TSHIRT_LP3.PNG	0.0533
GIRL_LM3.PNG	1.2574
GIRL_LM5.PNG	2.0413
GIRL_LP3.PNG	1.2574
GIRL_LP5.PNG	2.0434
GIRL_REGION_BACK_LM3.PNG	0.5209
GIRL_REGION_BACK_LP3.PNG	0.5209
GIRL_REGION_FRONT_LM3.PNG	0.7364
GIRL_REGION_FRONT_LP3.PNG	0.7365
JP_LM3.PNG	0.9973
JP_LM5.PNG	1.7318
JP_LP3.PNG	0.9903
JP_LP5.PNG	1.6033
JP_REGION_STRIPE2_LM3.PNG	0.0820
JP_REGION_STRIPE2_LP3.PNG	0.0830
JP_REGION_STRIPE_LM3.PNG	0.1707
JP_REGION_STRIPE_LP3.PNG	0.1721
TORE_LM3.PNG	1.0096
TORE_LM5.PNG	1.6790
TORE_LP3.PNG	0.9001
TORE_LP5.PNG	1.4530
TORE_REGION_BACK_LM3.PNG	0.3075
TORE_REGION_BACK_LP3.PNG	0.2728
TORE_REGION_FRONT_LM3.PNG	0.7021
TORE_REGION_FRONT_LP3.PNG	0.6273

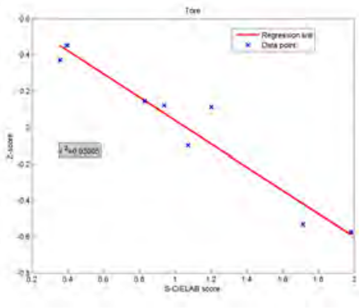
Table 54: Combined eye tracker and observer regions: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.55133
CARTOON_LM5.PNG	0.54835
CARTOON_LP3.PNG	0.55183
CARTOON_LP5.PNG	0.54956
CARTOON_REGION_SIGN_LM3.PNG	0.55239
CARTOON_REGION_SIGN_LP3.PNG	0.55459
CARTOON_REGION_TSHIRT_LM3.PNG	0.55287
CARTOON_REGION_TSHIRT_LP3.PNG	0.55281
GIRL_LM3.PNG	0.40390
GIRL_LM5.PNG	0.39927
GIRL_LP3.PNG	0.40474
GIRL_LP5.PNG	0.40309
GIRL_REGION_BACK_LM3.PNG	0.40437
GIRL_REGION_BACK_LP3.PNG	0.40467
GIRL_REGION_FRONT_LM3.PNG	0.40315
GIRL_REGION_FRONT_LP3.PNG	0.40367
JP_LM3.PNG	0.30006
JP_LM5.PNG	0.28672
JP_LP3.PNG	0.30386
JP_LP5.PNG	0.30112
JP_REGION_STRIPE2_LM3.PNG	0.32071
JP_REGION_STRIPE2_LP3.PNG	0.32212
JP_REGION_STRIPE_LM3.PNG	0.31731
JP_REGION_STRIPE_LP3.PNG	0.31755
TORE_LM3.PNG	0.29001
TORE_LM5.PNG	0.28716
TORE_LP3.PNG	0.29055
TORE_LP5.PNG	0.28541
TORE_REGION_BACK_LM3.PNG	0.29304
TORE_REGION_BACK_LP3.PNG	0.29353
TORE_REGION_FRONT_LM3.PNG	0.29030
TORE_REGION_FRONT_LP3.PNG	0.29048

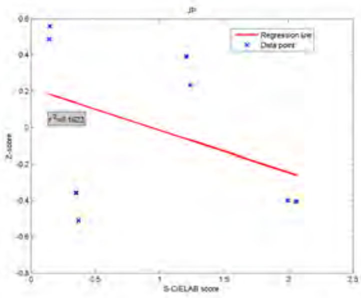
Table 55: Combined eye tracker and observer regions: SSIM score for all scenes and reproductions.



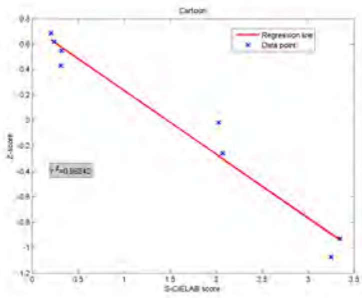
(a) Girl.



(b) Tore.

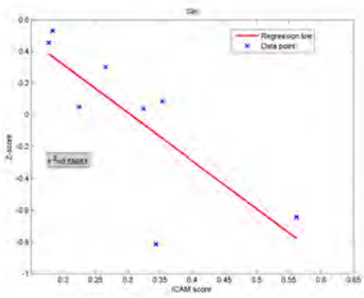


(c) JP.

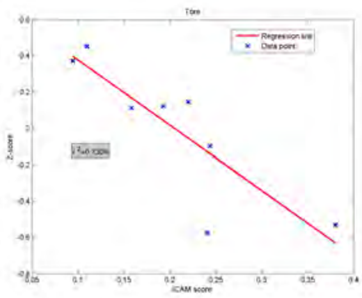


(d) Cartoon.

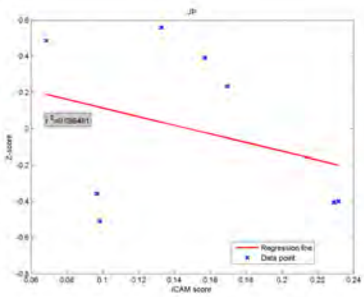
Figure 206: Combined eye tracker and observer regions: S-CIELAB regression plot for each scene.



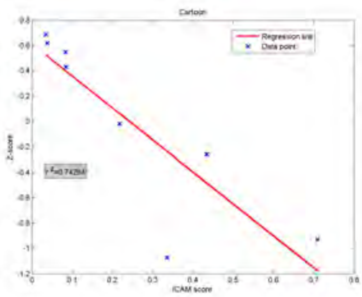
(a) Girl.



(b) Tore.

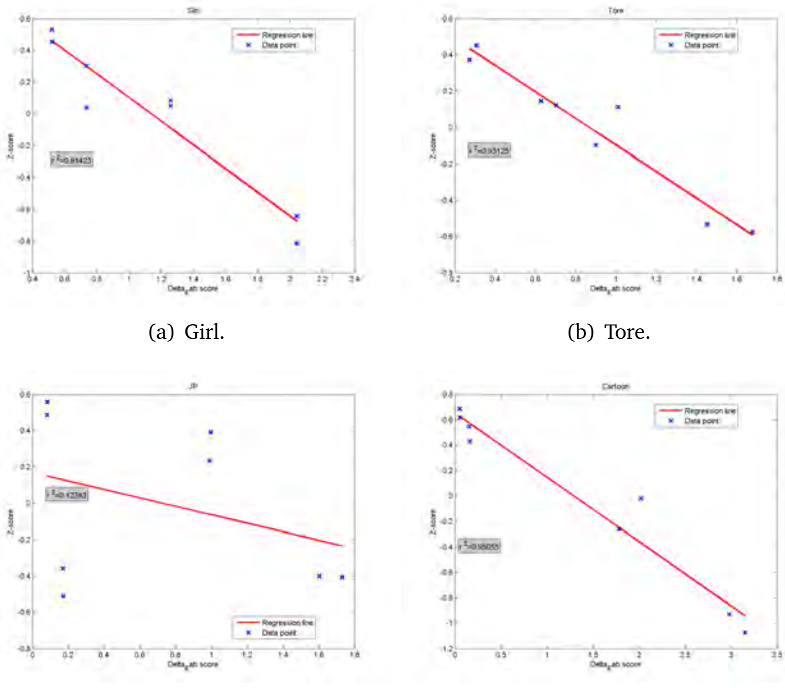


(c) JP.



(d) Cartoon.

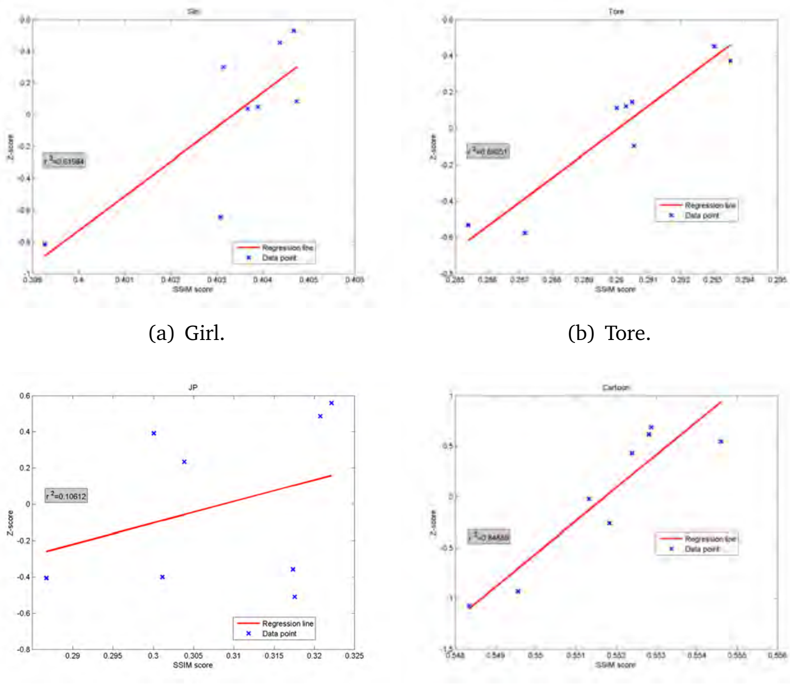
Figure 207: Combined eye tracker and observer regions: iCAM regression plot for each scene.



(a) Girl. (b) Tore.

(c) JP. (d) Cartoon.

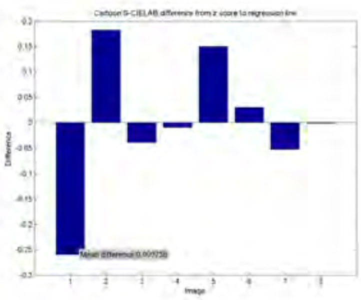
Figure 208: Combined eye tracker and observer regions: ΔE_{ab} regression plot for each scene.



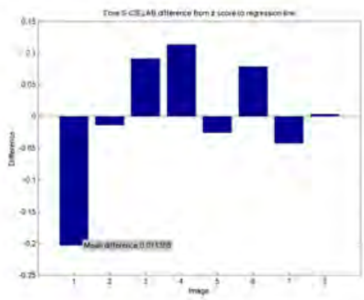
(a) Girl. (b) Tore.

(c) JP. (d) Cartoon.

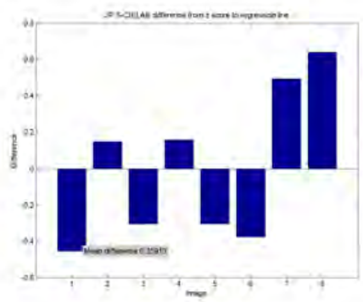
Figure 209: Combined eye tracker and observer regions: SSIM regression plot for each scene.



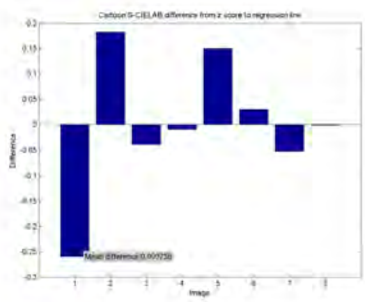
(a) Girl.



(b) Tore.

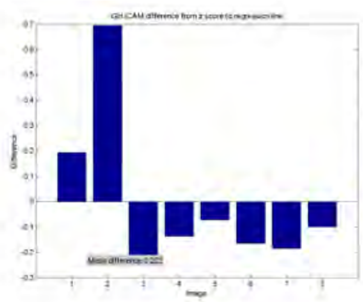


(c) JP.

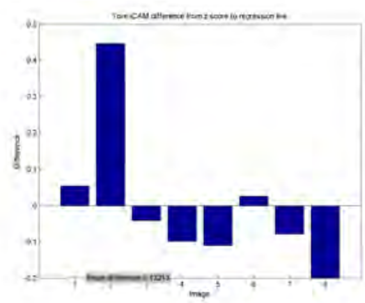


(d) Cartoon.

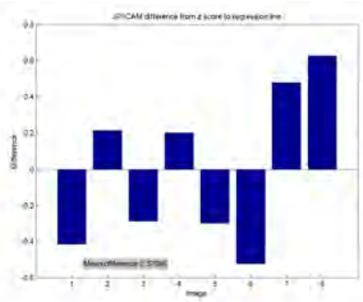
Figure 210: Combined eye tracker and observer regions: Difference from the regression line - S-CIELAB.



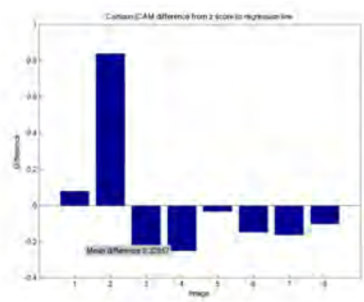
(a) Girl.



(b) Tore.

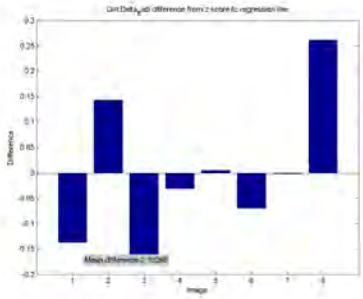


(c) JP.

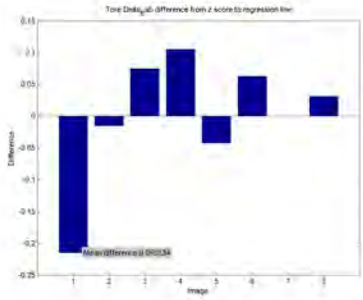


(d) Cartoon.

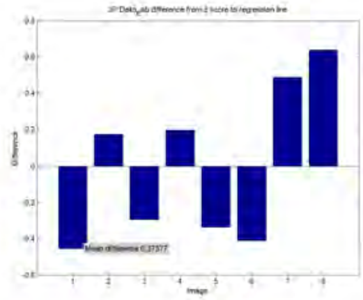
Figure 211: Combined eye tracker and observer regions: Difference from the regression line - iCAM.



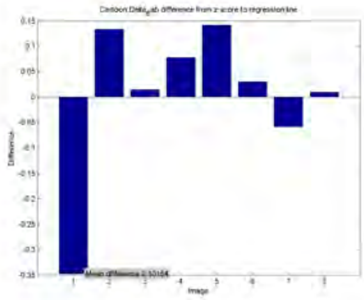
(a) Girl.



(b) Tore.

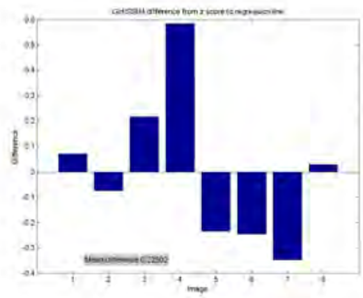


(c) JP.

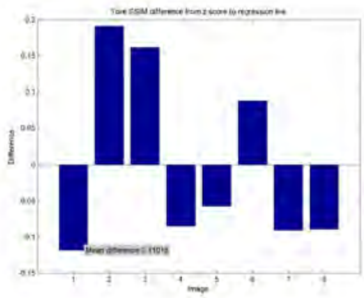


(d) Cartoon.

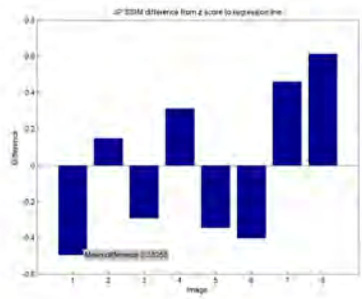
Figure 212: Combined eye tracker and observer regions: Difference from the regression line - ΔE_{ab}.



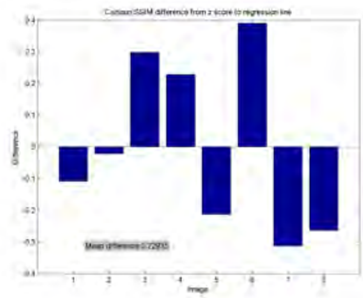
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 213: Combined eye tracker and observer regions: Difference from the regression line - SSIM.

D.6 Image difference metrics applied freeview regions

Scene	Mean
CARTOON_LM3.PNG	0.592
CARTOON_LM5.PNG	0.940
CARTOON_LP3.PNG	0.611
CARTOON_LP5.PNG	0.984
CARTOON_REGION_SIGN_LM3.PNG	0.139
CARTOON_REGION_SIGN_LP3.PNG	0.142
CARTOON_REGION_TSHIRT_LM3.PNG	0.089
CARTOON_REGION_TSHIRT_LP3.PNG	0.102
GIRL_LM3.PNG	0.298
GIRL_LM5.PNG	0.475
GIRL_LP3.PNG	0.303
GIRL_LP5.PNG	0.487
GIRL_REGION_BACK_LM3.PNG	0.040
GIRL_REGION_BACK_LP3.PNG	0.041
GIRL_REGION_FRONT_LM3.PNG	0.290
GIRL_REGION_FRONT_LP3.PNG	0.297
JP_LM3.PNG	0.205
JP_LM5.PNG	0.351
JP_LP3.PNG	0.212
JP_LP5.PNG	0.343
JP_REGION_STRIPE2_LM3.PNG	0.036
JP_REGION_STRIPE2_LP3.PNG	0.038
JP_REGION_STRIPE_LM3.PNG	0.008
JP_REGION_STRIPE_LP3.PNG	0.009
TORE_LM3.PNG	0.496
TORE_LM5.PNG	0.812
TORE_LP3.PNG	0.447
TORE_LP5.PNG	0.717
TORE_REGION_BACK_LM3.PNG	0.052
TORE_REGION_BACK_LP3.PNG	0.048
TORE_REGION_FRONT_LM3.PNG	0.469
TORE_REGION_FRONT_LP3.PNG	0.422

Table 56: Freeview regions: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.064
CARTOON_LM5.PNG	0.098
CARTOON_LP3.PNG	0.126
CARTOON_LP5.PNG	0.204
CARTOON_REGION_SIGN_LM3.PNG	0.034
CARTOON_REGION_SIGN_LP3.PNG	0.033
CARTOON_REGION_TSHIRT_LM3.PNG	0.015
CARTOON_REGION_TSHIRT_LP3.PNG	0.016
GIRL_LM3.PNG	0.049
GIRL_LM5.PNG	0.074
GIRL_LP3.PNG	0.069
GIRL_LP5.PNG	0.108
GIRL_REGION_BACK_LM3.PNG	0.024
GIRL_REGION_BACK_LP3.PNG	0.025
GIRL_REGION_FRONT_LM3.PNG	0.063
GIRL_REGION_FRONT_LP3.PNG	0.081
JP_LM3.PNG	0.036
JP_LM5.PNG	0.053
JP_LP3.PNG	0.040
JP_LP5.PNG	0.053
JP_REGION_STRIPE2_LM3.PNG	0.016
JP_REGION_STRIPE2_LP3.PNG	0.027
JP_REGION_STRIPE_LM3.PNG	0.006
JP_REGION_STRIPE_LP3.PNG	0.007
TORE_LM3.PNG	0.075
TORE_LM5.PNG	0.108
TORE_LP3.PNG	0.114
TORE_LP5.PNG	0.172
TORE_REGION_BACK_LM3.PNG	0.021
TORE_REGION_BACK_LP3.PNG	0.019
TORE_REGION_FRONT_LM3.PNG	0.083
TORE_REGION_FRONT_LP3.PNG	0.115

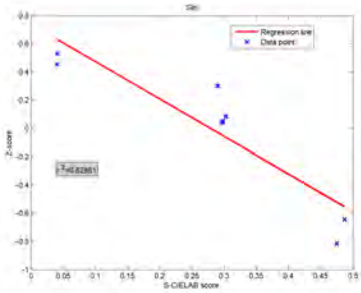
Table 57: Freeview regions: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.540
CARTOON_LM5.PNG	0.851
CARTOON_LP3.PNG	0.495
CARTOON_LP5.PNG	0.818
CARTOON_REGION_SIGN_LM3.PNG	0.071
CARTOON_REGION_SIGN_LP3.PNG	0.067
CARTOON_REGION_TSHIRT_LM3.PNG	0.023
CARTOON_REGION_TSHIRT_LP3.PNG	0.024
GIRL_LM3.PNG	0.219
GIRL_LM5.PNG	0.355
GIRL_LP3.PNG	0.219
GIRL_LP5.PNG	0.355
GIRL_REGION_BACK_LM3.PNG	0.013
GIRL_REGION_BACK_LP3.PNG	0.013
GIRL_REGION_FRONT_LM3.PNG	0.206
GIRL_REGION_FRONT_LP3.PNG	0.206
JP_LM3.PNG	0.177
JP_LM5.PNG	0.307
JP_LP3.PNG	0.178
JP_LP5.PNG	0.289
JP_REGION_STRIPE2_LM3.PNG	0.024
JP_REGION_STRIPE2_LP3.PNG	0.025
JP_REGION_STRIPE_LM3.PNG	0.003
JP_REGION_STRIPE_LP3.PNG	0.003
TORE_LM3.PNG	0.391
TORE_LM5.PNG	0.647
TORE_LP3.PNG	0.359
TORE_LP5.PNG	0.582
TORE_REGION_BACK_LM3.PNG	0.035
TORE_REGION_BACK_LP3.PNG	0.031
TORE_REGION_FRONT_LM3.PNG	0.356
TORE_REGION_FRONT_LP3.PNG	0.328

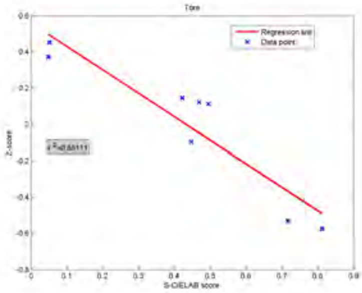
Table 58: Freeview regions: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.159
CARTOON_LM5.PNG	0.158
CARTOON_LP3.PNG	0.159
CARTOON_LP5.PNG	0.158
CARTOON_REGION_SIGN_LM3.PNG	0.160
CARTOON_REGION_SIGN_LP3.PNG	0.161
CARTOON_REGION_TSHIRT_LM3.PNG	0.160
CARTOON_REGION_TSHIRT_LP3.PNG	0.160
GIRL_LM3.PNG	0.072
GIRL_LM5.PNG	0.071
GIRL_LP3.PNG	0.072
GIRL_LP5.PNG	0.072
GIRL_REGION_BACK_LM3.PNG	0.072
GIRL_REGION_BACK_LP3.PNG	0.072
GIRL_REGION_FRONT_LM3.PNG	0.072
GIRL_REGION_FRONT_LP3.PNG	0.072
JP_LM3.PNG	0.053
JP_LM5.PNG	0.051
JP_LP3.PNG	0.055
JP_LP5.PNG	0.054
JP_REGION_STRIPE2_LM3.PNG	0.058
JP_REGION_STRIPE2_LP3.PNG	0.058
JP_REGION_STRIPE_LM3.PNG	0.058
JP_REGION_STRIPE_LP3.PNG	0.058
TORE_LM3.PNG	0.114
TORE_LM5.PNG	0.113
TORE_LP3.PNG	0.116
TORE_LP5.PNG	0.115
TORE_REGION_BACK_LM3.PNG	0.119
TORE_REGION_BACK_LP3.PNG	0.119
TORE_REGION_FRONT_LM3.PNG	0.114
TORE_REGION_FRONT_LP3.PNG	0.116

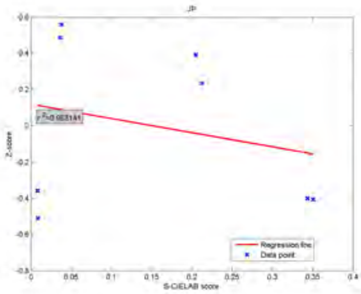
Table 59: Freeview regions: SSIM score for all scenes and reproductions.



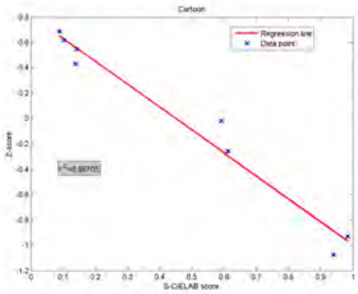
(a) Girl.



(b) Tore.

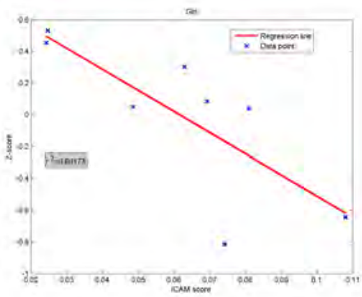


(c) JP.

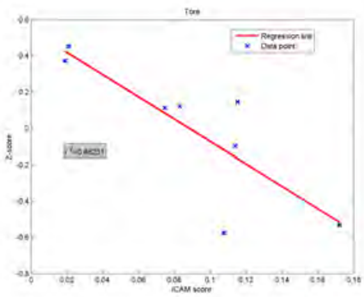


(d) Cartoon.

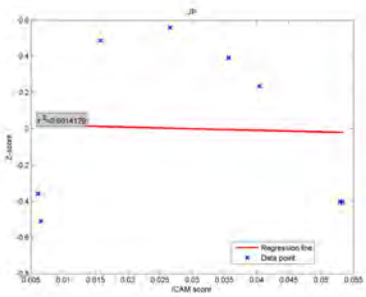
Figure 214: Freeview regions: S-CIELAB regression plot for each scene.



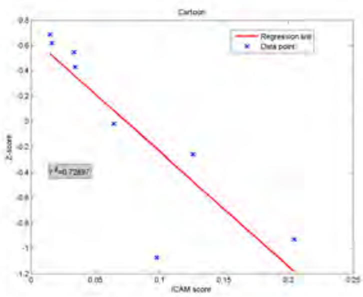
(a) Girl.



(b) Tore.

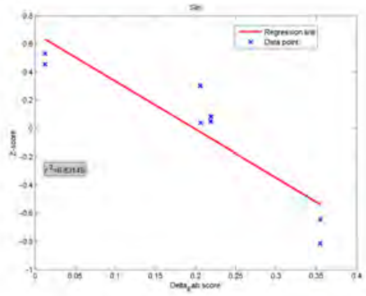


(c) JP.

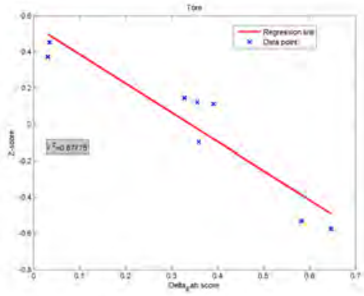


(d) Cartoon.

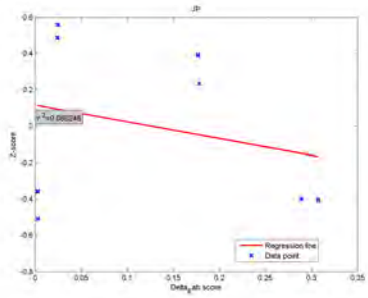
Figure 215: Freeview regions: iCAM regression plot for each scene.



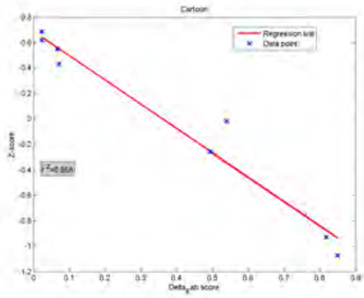
(a) Girl.



(b) Tore.

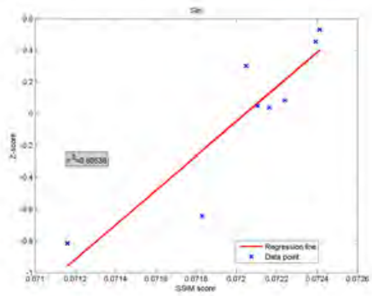


(c) JP.

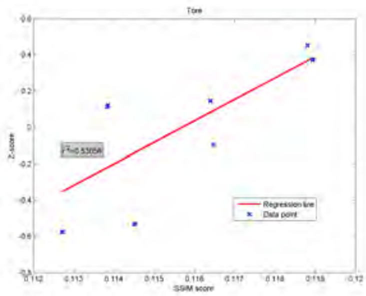


(d) Cartoon.

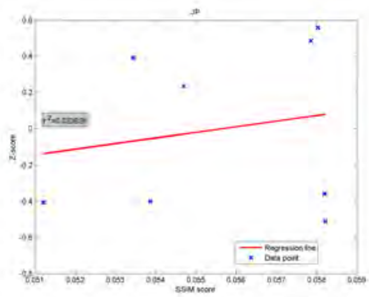
Figure 216: Freeview regions: ΔE_{ab} regression plot for each scene.



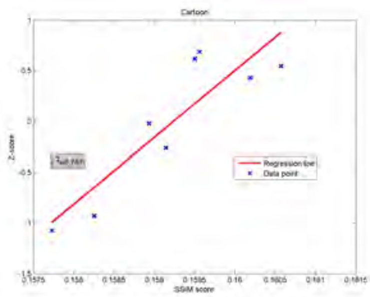
(a) Girl.



(b) Tore.

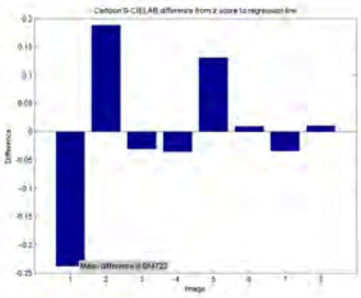


(c) JP.

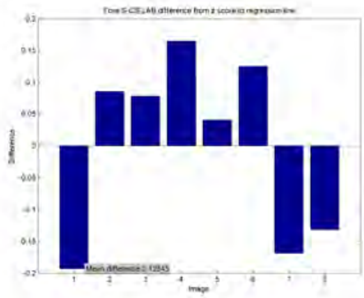


(d) Cartoon.

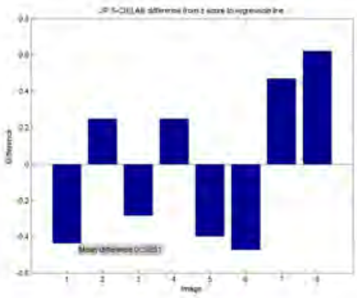
Figure 217: Freeview regions: SSIM regression plot for each scene.



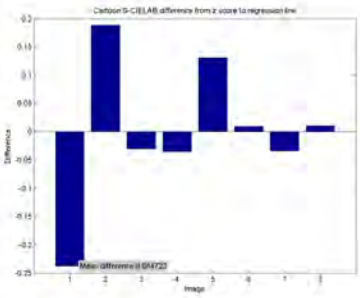
(a) Girl.



(b) Tore.

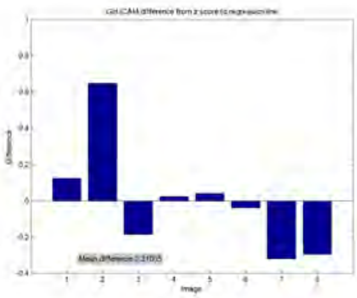


(c) JP.

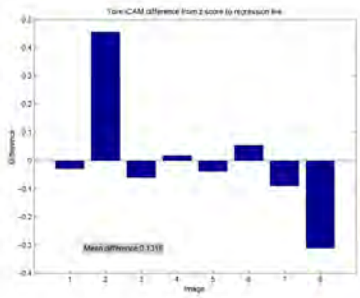


(d) Cartoon.

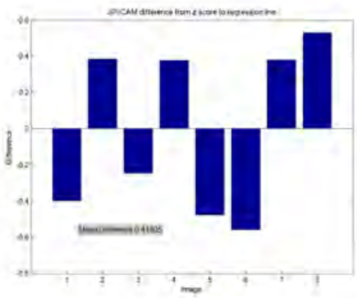
Figure 218: Freeview regions: Difference from the regression line - S-CIELAB.



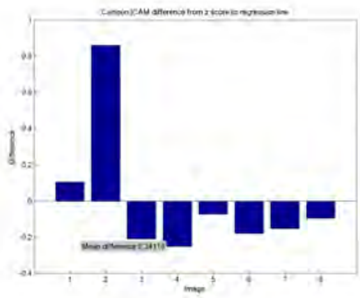
(a) Girl.



(b) Tore.

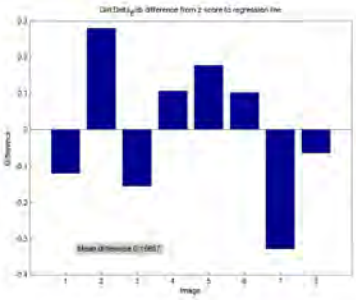


(c) JP.

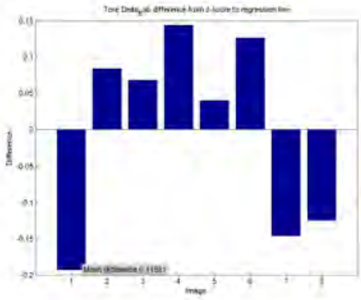


(d) Cartoon.

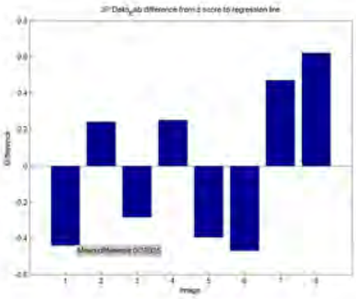
Figure 219: Freeview regions: Difference from the regression line - iCAM.



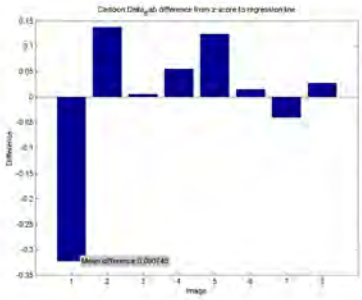
(a) Girl.



(b) Tore.

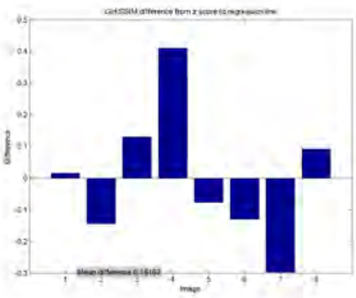


(c) JP.

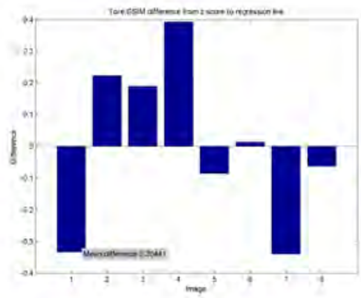


(d) Cartoon.

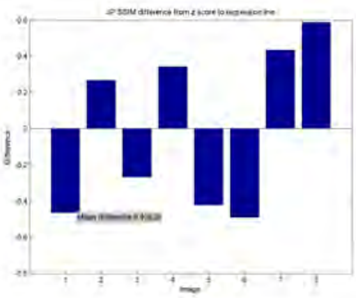
Figure 220: Freeview regions: Difference from the regression line - ΔE_{ab}.



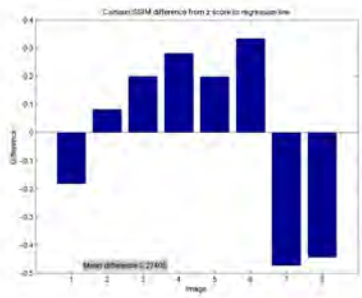
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 221: Freeview regions: Difference from the regression line - SSIM.

D.7 Image difference metrics applied gaze marking regions

Scene	Mean
CARTOON_LM3.PNG	0.555
CARTOON_LM5.PNG	0.883
CARTOON_LP3.PNG	0.569
CARTOON_LP5.PNG	0.918
CARTOON_REGION_SIGN_LM3.PNG	0.080
CARTOON_REGION_SIGN_LP3.PNG	0.082
CARTOON_REGION_TSHIRT_LM3.PNG	0.073
CARTOON_REGION_TSHIRT_LP3.PNG	0.083
GIRL_LM3.PNG	0.615
GIRL_LM5.PNG	0.983
GIRL_LP3.PNG	0.624
GIRL_LP5.PNG	1.006
GIRL_REGION_BACK_LM3.PNG	0.153
GIRL_REGION_BACK_LP3.PNG	0.155
GIRL_REGION_FRONT_LM3.PNG	0.553
GIRL_REGION_FRONT_LP3.PNG	0.563
JP_LM3.PNG	0.449
JP_LM5.PNG	0.768
JP_LP3.PNG	0.465
JP_LP5.PNG	0.753
JP_REGION_STRIPE2_LM3.PNG	0.060
JP_REGION_STRIPE2_LP3.PNG	0.062
JP_REGION_STRIPE_LM3.PNG	0.034
JP_REGION_STRIPE_LP3.PNG	0.036
TORE_LM3.PNG	0.490
TORE_LM5.PNG	0.800
TORE_LP3.PNG	0.452
TORE_LP5.PNG	0.725
TORE_REGION_BACK_LM3.PNG	0.104
TORE_REGION_BACK_LP3.PNG	0.101
TORE_REGION_FRONT_LM3.PNG	0.426
TORE_REGION_FRONT_LP3.PNG	0.390

Table 60: Gaze marking regions: S-CIELAB score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.061
CARTOON_LM5.PNG	0.094
CARTOON_LP3.PNG	0.122
CARTOON_LP5.PNG	0.198
CARTOON_REGION_SIGN_LM3.PNG	0.023
CARTOON_REGION_SIGN_LP3.PNG	0.023
CARTOON_REGION_TSHIRT_LM3.PNG	0.012
CARTOON_REGION_TSHIRT_LP3.PNG	0.013
GIRL_LM3.PNG	0.101
GIRL_LM5.PNG	0.153
GIRL_LP3.PNG	0.147
GIRL_LP5.PNG	0.231
GIRL_REGION_BACK_LM3.PNG	0.060
GIRL_REGION_BACK_LP3.PNG	0.061
GIRL_REGION_FRONT_LM3.PNG	0.128
GIRL_REGION_FRONT_LP3.PNG	0.159
JP_LM3.PNG	0.103
JP_LM5.PNG	0.160
JP_LP3.PNG	0.115
JP_LP5.PNG	0.144
JP_REGION_STRIPE2_LM3.PNG	0.033
JP_REGION_STRIPE2_LP3.PNG	0.060
JP_REGION_STRIPE_LM3.PNG	0.018
JP_REGION_STRIPE_LP3.PNG	0.019
TORE_LM3.PNG	0.095
TORE_LM5.PNG	0.136
TORE_LP3.PNG	0.143
TORE_LP5.PNG	0.212
TORE_REGION_BACK_LM3.PNG	0.040
TORE_REGION_BACK_LP3.PNG	0.037
TORE_REGION_FRONT_LM3.PNG	0.100
TORE_REGION_FRONT_LP3.PNG	0.136

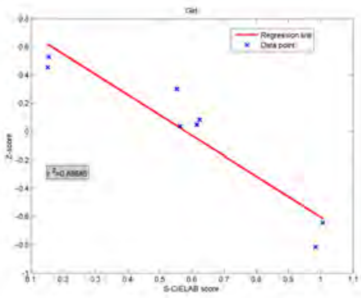
Table 61: Gaze marking regions: iCAM score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.531
CARTOON_LM5.PNG	0.834
CARTOON_LP3.PNG	0.481
CARTOON_LP5.PNG	0.798
CARTOON_REGION_SIGN_LM3.PNG	0.043
CARTOON_REGION_SIGN_LP3.PNG	0.041
CARTOON_REGION_TSHIRT_LM3.PNG	0.019
CARTOON_REGION_TSHIRT_LP3.PNG	0.020
GIRL_LM3.PNG	0.487
GIRL_LM5.PNG	0.790
GIRL_LP3.PNG	0.487
GIRL_LP5.PNG	0.791
GIRL_REGION_BACK_LM3.PNG	0.088
GIRL_REGION_BACK_LP3.PNG	0.088
GIRL_REGION_FRONT_LM3.PNG	0.398
GIRL_REGION_FRONT_LP3.PNG	0.399
JP_LM3.PNG	0.417
JP_LM5.PNG	0.720
JP_LP3.PNG	0.412
JP_LP5.PNG	0.667
JP_REGION_STRIPE2_LM3.PNG	0.040
JP_REGION_STRIPE2_LP3.PNG	0.040
JP_REGION_STRIPE_LM3.PNG	0.014
JP_REGION_STRIPE_LP3.PNG	0.014
TORE_LM3.PNG	0.409
TORE_LM5.PNG	0.674
TORE_LP3.PNG	0.383
TORE_LP5.PNG	0.619
TORE_REGION_BACK_LM3.PNG	0.085
TORE_REGION_BACK_LP3.PNG	0.074
TORE_REGION_FRONT_LM3.PNG	0.323
TORE_REGION_FRONT_LP3.PNG	0.309

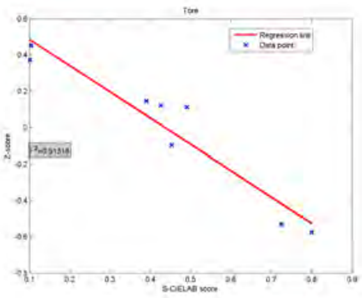
Table 62: Gaze marking regions: ΔE_{ab} score for all scenes and reproductions.

Scene	Mean
CARTOON_LM3.PNG	0.153
CARTOON_LM5.PNG	0.152
CARTOON_LP3.PNG	0.153
CARTOON_LP5.PNG	0.153
CARTOON_REGION_SIGN_LM3.PNG	0.154
CARTOON_REGION_SIGN_LP3.PNG	0.155
CARTOON_REGION_TSHIRT_LM3.PNG	0.154
CARTOON_REGION_TSHIRT_LP3.PNG	0.154
GIRL_LM3.PNG	0.161
GIRL_LM5.PNG	0.159
GIRL_LP3.PNG	0.161
GIRL_LP5.PNG	0.160
GIRL_REGION_BACK_LM3.PNG	0.161
GIRL_REGION_BACK_LP3.PNG	0.161
GIRL_REGION_FRONT_LM3.PNG	0.160
GIRL_REGION_FRONT_LP3.PNG	0.161
JP_LM3.PNG	0.115
JP_LM5.PNG	0.106
JP_LP3.PNG	0.120
JP_LP5.PNG	0.117
JP_REGION_STRIPE2_LM3.PNG	0.133
JP_REGION_STRIPE2_LP3.PNG	0.133
JP_REGION_STRIPE_LM3.PNG	0.133
JP_REGION_STRIPE_LP3.PNG	0.133
TORE_LM3.PNG	0.117
TORE_LM5.PNG	0.115
TORE_LP3.PNG	0.122
TORE_LP5.PNG	0.119
TORE_REGION_BACK_LM3.PNG	0.125
TORE_REGION_BACK_LP3.PNG	0.127
TORE_REGION_FRONT_LM3.PNG	0.118
TORE_REGION_FRONT_LP3.PNG	0.122

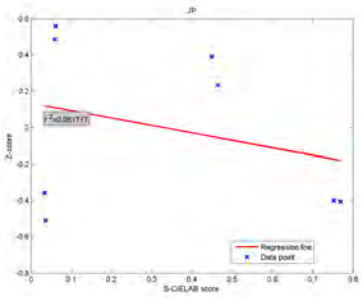
Table 63: Gaze marking regions: SSIM score for all scenes and reproductions.



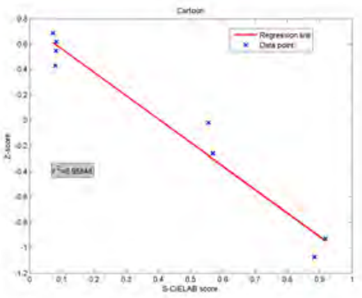
(a) Girl.



(b) Tore.

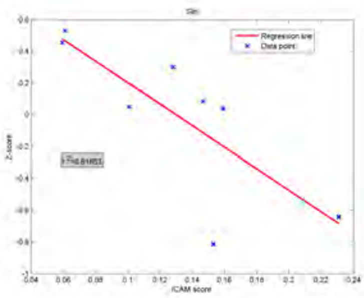


(c) JP.

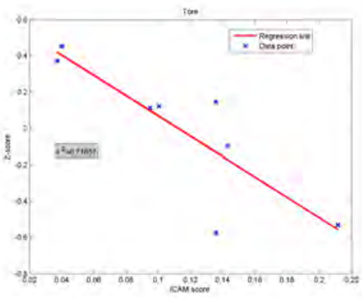


(d) Cartoon.

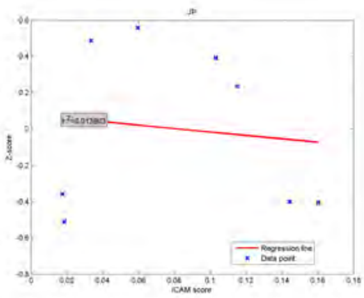
Figure 222: Gaze marking regions: S-CIELAB regression plot for each scene.



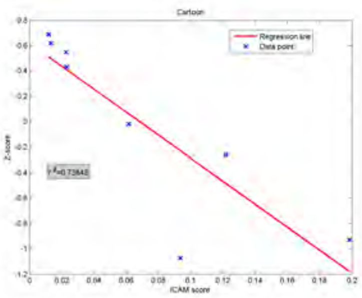
(a) Girl.



(b) Tore.

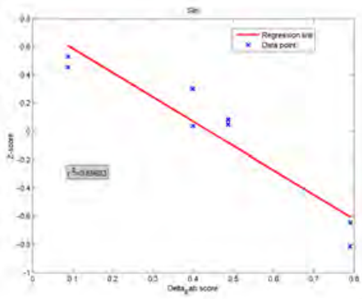


(c) JP.

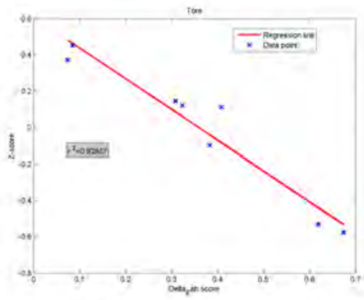


(d) Cartoon.

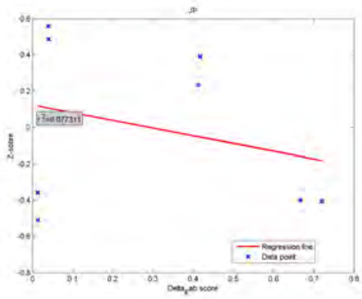
Figure 223: Gaze marking regions: iCAM regression plot for each scene.



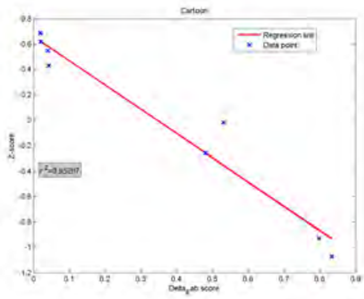
(a) Girl.



(b) Tore.

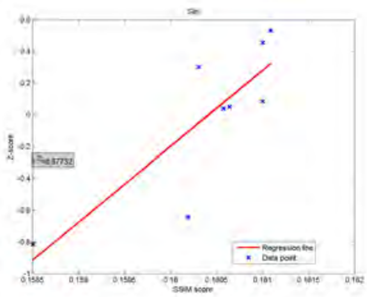


(c) JP.

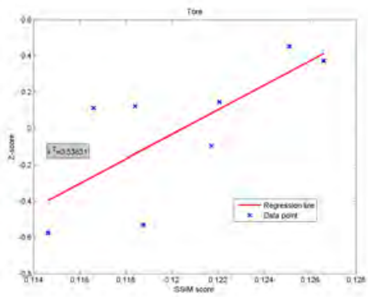


(d) Cartoon.

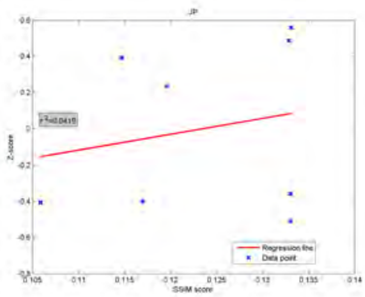
Figure 224: Gaze marking regions: ΔE_{ab} regression plot for each scene.



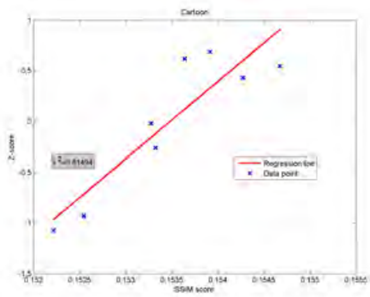
(a) Girl.



(b) Tore.

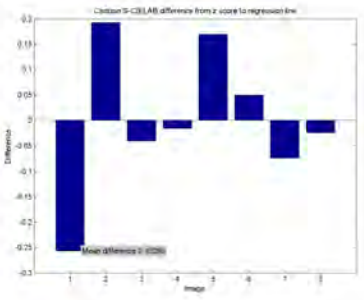


(c) JP.

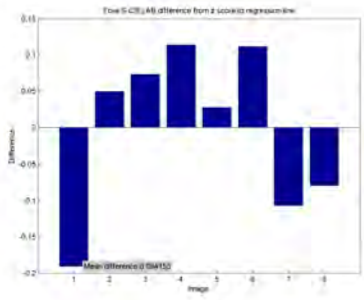


(d) Cartoon.

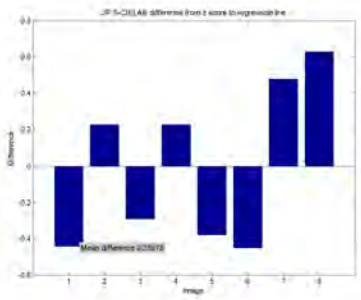
Figure 225: Gaze marking regions: SSIM regression plot for each scene.



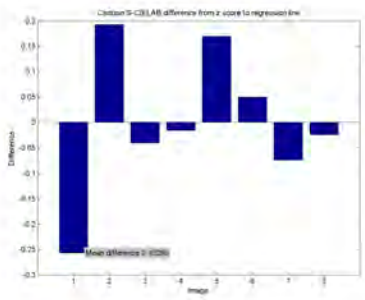
(a) Girl.



(b) Tore.

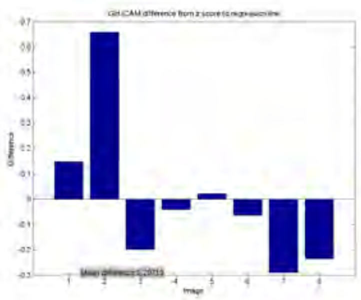


(c) JP.

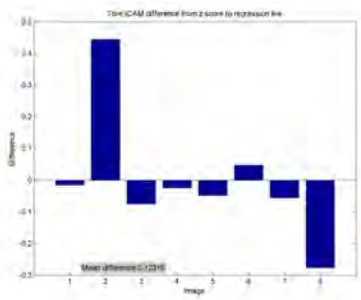


(d) Cartoon

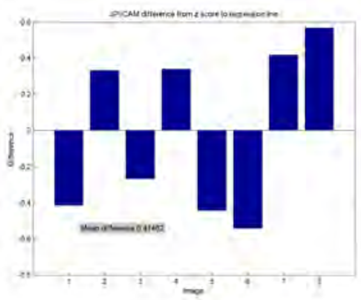
Figure 226: Gaze marking regions: Difference from the regression line - S-CIELAB.



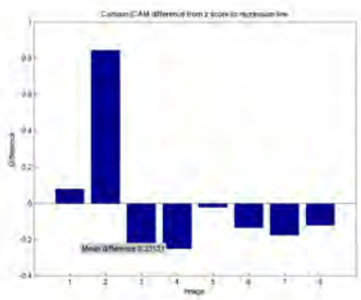
(a) Girl.



(b) Tore.

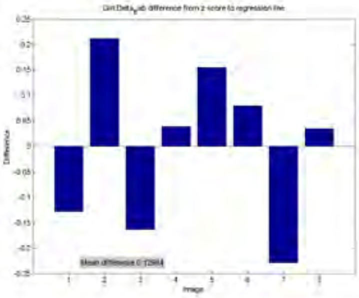


(c) JP.

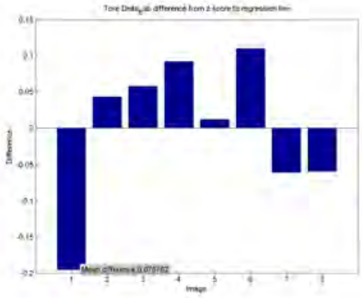


(d) Cartoon.

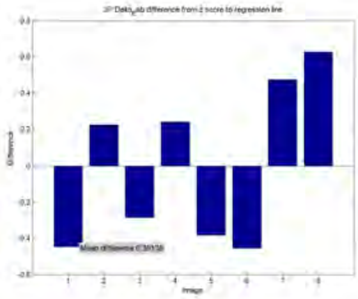
Figure 227: Gaze marking regions: Difference from the regression line - iCAM.



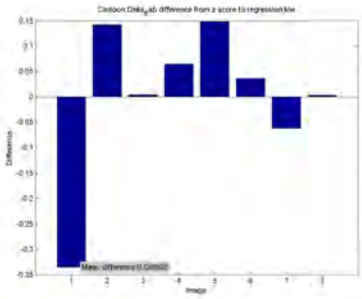
(a) Girl.



(b) Tore.

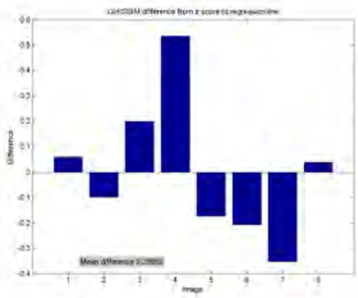


(c) JP.

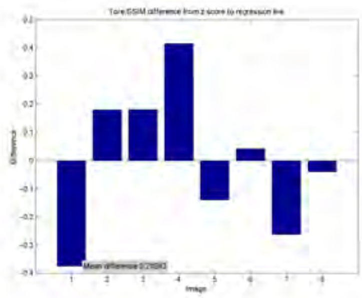


(d) Cartoon.

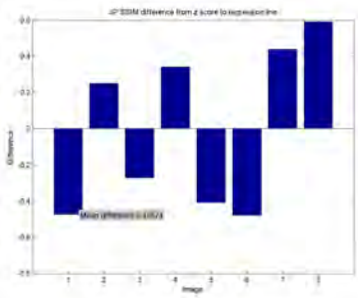
Figure 228: Gaze marking regions: Difference from the regression line - ΔE_{ab}.



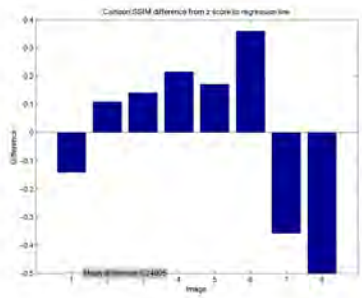
(a) Girl.



(b) Tore.



(c) JP.



(d) Cartoon.

Figure 229: Gaze marking regions: Difference from the regression line - SSIM.