Visual Attention Retargeting

by

Victor A. Mateescu

B.A.Sc., Simon Fraser University, 2012

Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Applied Science

in the

School of Engineering Science

Faculty of Applied Sciences

APPROVAL

Name: Victor A. Mateescu

Degree: Master of Applied Science

Title: Visual Attention Retargeting

Examining Committee: **Chair**: **Dr. Rodney G. Vaughan**

Professor, School of Engineering Science

_____

Dr. Ivan V. Bajić

Senior Supervisor

Associate Professor, School of Engineering Science

_____

Dr. Parvaneh Saeedi

Supervisor

Associate Professor, School of Engineering Science

_____

Dr. Jie Liang

Internal Examiner

Associate Professor, School of Engineering Science

Date Approved: December 8th, 2014

Partial Copyright Licence

SFU

Abstract

This thesis explores attention retargeting—a concept related to visual saliency where the content or composition of an image is altered in an effort to guide the viewer's attention. Attention retargeting is currently in its infancy with numerous unexplored possibilities, no common methodology for evaluating performance, and no unified framework. The difficulty of attention retargeting as a saliency inversion problem lies in the lack of one-to-one mapping between saliency and the image domain, in addition to the possible negative impact of saliency alterations on image naturalness. Several approaches from recent literature to solve this challenging problem are reviewed in this context. Two novel attention retargeting methods are proposed to efficiently compute a region's propensity for drawing attention after it has been modified. The first method manipulates the orientation of a selected region, while the second modifies its hue. Both methods are applied to maximize the saliency of selected regions in various images. The likelihood of drawing attention towards the modified regions is evaluated through eye-tracking. Subjective experiments, in which participants are told to decide which image looks better between two alternatives, are used to measure the relative naturalness of the modification. An experiment was conducted to determine whether subliminal flicker is capable of drawing attention in natural images without the viewer's knowledge. Flicker was introduced to selected regions in a set of images by alternating the contrast in these regions from high to low at a frequency of 50 Hz. A comparison of eye-tracking data between participants who viewed the flickering images against those who viewed the original images suggests that subliminal flicker may, on average, repel attention rather than attract it.

Keywords: attention retargeting; visual saliency; subliminal cues; image processing

To my mother and father for supporting me in all my endeavors.
To my sister for hogging all the artistic talent in the family,
thus driving me to my current passion.

Acknowledgements

The debt I owe to my supervisor Prof. Ivan V. Bajić for his expert guidance is incalculable. My gratitude towards him reaches beyond his contributions to my work. I'd like to thank him for instilling his sense of integrity in me, for teaching me to think outside the box (admittedly, I occasionally mistook his creative research ideas as sporadic lapses in sanity), and for his strong work ethic that I strive to emulate. His role as my mentor these past two-and-a-half years will not be forgotten.

I am very grateful to Dr. Hadi Hadizadeh for laying the foundation for my work. His assistance in setting up the eye-tracking equipment and software while working a full-time job is much appreciated. His motivation and support as I took those first intimidating steps into graduate school were invaluable.

My sincere thanks to Dr. Mario J. Enriquez for his generous aid during the harrowing weeks when our eye-tracking equipment had malfunctioned. The work presented in this thesis could not have come to full fruition without his critical repairs.

Last but not least, I'd like to thank my fellow lab mates Mehdi Stapleton, Hossein Khatoonabadi, and Hanieh Khalilian for fighting the urge to sleep as I presented my work during our group meetings. I am grateful for their thoughtful questions and the feedback and support they provided.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Attention retargeting is a newly emerging field of research in which the content of an image or video is altered in an effort to guide a viewer's attention. The origins of attention retargeting are closely tied to visual saliency, a topic that addresses the question "where do we look?" As electronic devices steadily improve, remarkable developments in multimedia, such as ultra high definition television and high framerate video, are becoming commonplace. Whether we recognize it or not, our eyes are spoiled beyond belief. With the emergence of these technologies and the progress made in saliency research, a new question—"where should we look?"—has become equally relevant.

The most immediately obvious application of attention retargeting is in advertising. A noteworthy example would be a large-scale application to the YouTube thumbnails belonging to a user's videos to help draw viewers. Another possible application is in visualization tools. Certain aspects in the creation of complex plots and diagrams could be automated, such as color selection, line styles, and overall layout, to implicitly guide the viewer and help improve clarity. Attention retargeting may also be applied to error concealment in video, where attention-grabbing artifacts are masked to improve subjective quality. Image summarization techniques could also benefit from attention retargeting. For example, distracting regions could be de-emphasized to guide a viewer's attention along a specific trajectory. Other applications include education, gaming, design and media arts.

Attention retargeting is a relatively unexplored topic with few contributions. Some of the main issues associated with this topic are: 1) unexplored possibilities in terms of visual features modified to manipulate saliency; 2) the potentially damaging effects of saliency alterations on image naturalness; 3) high computational complexity inherited in part from

1

saliency computation; 4) no well-established measure of performance; 5) no unified framework. The collective work presented in this thesis addresses each of these issues. Since attention retargeting builds upon principles of early attention, we begin with a discussion of visual saliency. We describe attention retargeting as a saliency inversion problem and provide some general approaches to solving it based on existing work—a modest step towards a unified framework. This clarifies fundamental differences in the methodology of existing attention retargeting models.

1.1 Visual Saliency

A full, detailed analysis of every object within view at any single point in time is an incredibly complex task, far beyond the capabilities of human visual processing [37]. And yet, most would agree that the simple act of seeing is a rather effortless endeavor; sitting on a couch watching television is not often thought of as a taxing activity. To cope with the vast amount of input obtained through our eyes [16], our attentional mechanisms will focus on a few key areas for in-depth analysis using a two-stage process [36].

The initial processing narrows down the list of potentially relevant areas in the scene by identifying objects that appear to pop out in terms of basic visual features, such as color, orientation, and motion [38]. An interesting or conspicuous object (in the context of these visual primitives) can be perceived within 25 - 50 ms [16] in an effortless manner. The second stage of attention involves a more complex, detailed processing out of the viewer's own volition. Although the deliberate nature of this mechanism can override the attention given to the objects that popped out in the first stage, it cannot be deployed at such a fast rate. This means that if the scene suddenly changes, we involuntarily draw our gaze toward objects that are sufficiently salient in the pre-attentive stage regardless of any intent to avoid them. For example, you'll likely be distracted for a quick moment by the backlit screen of a cell phone in a dark movie theater the instant someone in front of you takes it out to start texting. Soon afterwards however, you resolve to ignore it and return your focus onto the movie.

Visual saliency primarily refers to the initial stage of visual attention. A region of interest (ROI) is salient if it is perceptually different from its local surroundings in terms of basic visual features. An example of this is illustrated in Fig. 1.1(a). Among the three

Figure 1.1: Illustration of visual saliency in terms of (a) orientation and (b) color.

circular regions, most would agree that the one on the bottom left is the most salient. This is because the bottom-left region is orthogonal, and hence, the most dissimilar to the vertical grating surrounding it, whereas the difference is far less pronounced for the other two regions. In an analogous example with color shown in Fig. 1.1(b), most would agree that the bottom-right circle is the most salient.

A popular architecture for visual saliency computation is illustrated in Fig. 1.2. This architecture was originally proposed in [17] and has been utilized in several other works [31, 21, 13, 11, 9]. A typical model first extracts a set of visual features that drive early attention from the input stimuli. In the subsequent stage, conspicuities are identified within each feature channel by computing local dissimilarity among features corresponding to each region of the input. For example, the image intensity map can be divided into non-overlapping square patches, each of which is compared against their adjacent patches. This operation, frequently referred to as "center-surround," produces a conspicuity map for each feature channel, which details the particular feature's influence on overall saliency in the original image. Each conspicuity map must be normalized to ensure that their values fall within a common range and to punish uniformity. Finally, the normalized conspicuity maps are linearly combined.

The final result is a grayscale image called a saliency map, as illustrated in Fig. 1.3 (top-right). Highly salient regions that are predicted to draw viewer's attention are denoted by pixels with high values (bright), whereas non-salient regions that are likely to be neglected are given low values (dark). The above architecture can be applied to predict saliency in videos as well. The input would be a sequence of frames rather than an image, and temporal

3
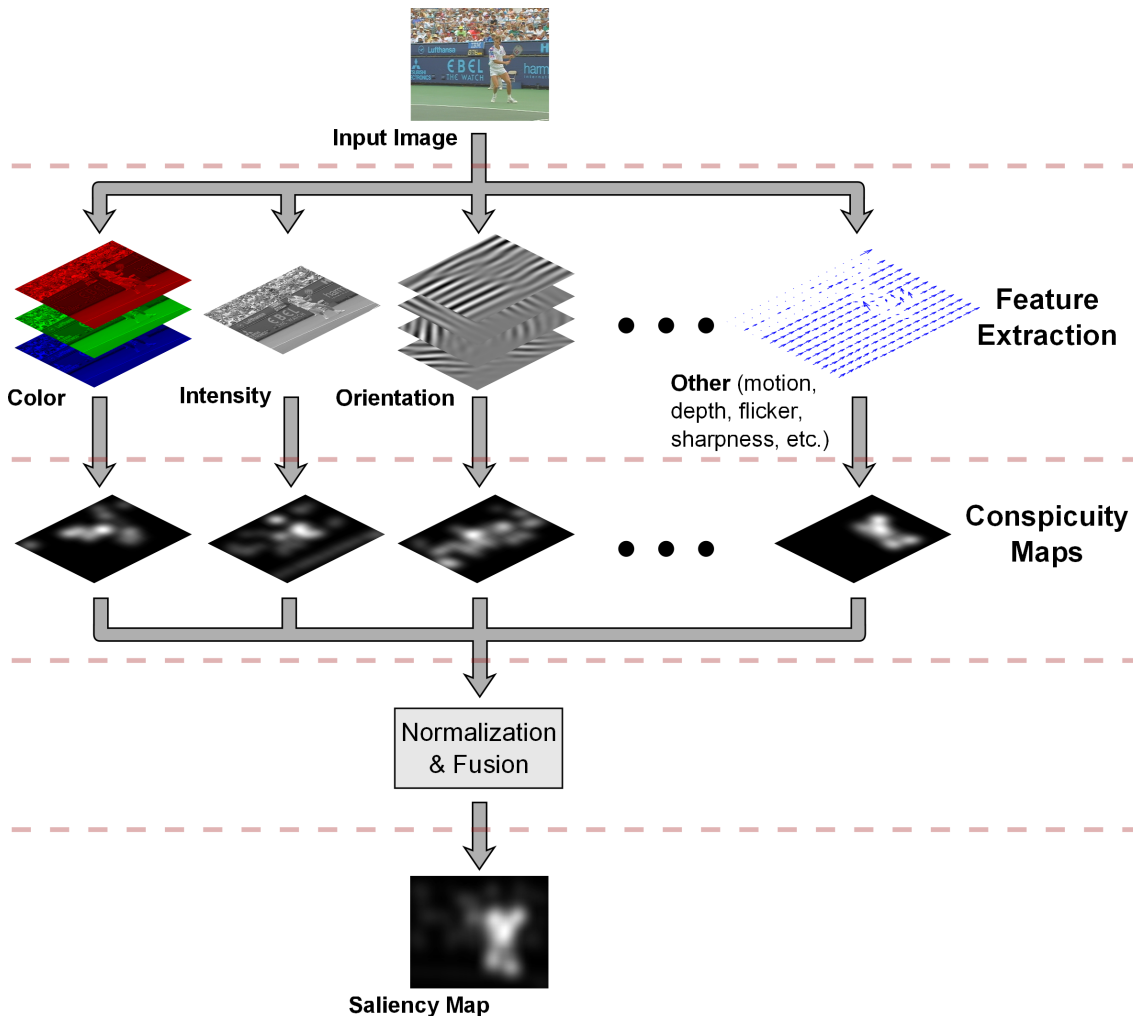
Figure 1.2: A popular model for visual saliency computation. (Diagram adapted from [17])

features, such as flicker and motion, would be utilized in addition to spatial features to produce a saliency map for each frame in the video.

Although the general problem of computational attention modeling remains largely unsolved, the past few decades of research demonstrate a fairly decent grasp of predicting visual saliency [1]. This is significant because the involuntary nature of visual saliency makes it a powerful tool for orienting attention. We may then pose an interesting problem: if we can compute visual saliency with reasonable accuracy, how can we alter it to manipulate a viewer's attention?

1.2 Attention Retargeting

Attention retargeting refers to modifying an image or video in an effort to alter viewer's gaze patterns in a desired way. This can be thought of as a saliency inversion problem, as illustrated in Fig. 1.3. Given an image and a map of desired saliency (called a target saliency map), we want to obtain the modified image that matches our desired saliency as the inverse of the target saliency map. Unfortunately, this is an ill-posed problem since there is no one-to-one mapping between saliency and images. This is largely due to the fact that saliency can stem from various different features. As illustrated in Fig. 1.3, the desired change in saliency can be achieved through the manipulation of features like intensity, color, or spatial frequency, among others. This problem can persist even when the retargeting is constrained to a single feature, since saliency is based on context. For example, the saliency of a ROI can be altered by raising the intensity of the ROI, or decreasing the intensity of its surroundings, or a combination of the two.

The fact that a desired change in saliency can be obtained in many ways leads to the impression that modifying saliency is easy. This is a difficult point to contest; a drastic change in the intensity of a region can easily make it stand out, and an application of Gaussian blurring to reduce visual conspicuities may serve to conceal it. Although modifying saliency may be simple in itself, doing so in a manner that preserves the naturalness of the original image is not. The concept of naturalness in images is fairly subjective and difficult to measure.

Attention retargeting is relatively unexplored and lacks a unified framework. We begin the following subsection with a general approach to this problem. Examples and ideas for improvement are provided with a brief overview of existing work in the field.

1.2.1 Iterative Black-Box Approach

The simplest approach to attention retargeting is to use saliency computation as a black box in an iterative optimization procedure, as shown in Fig. 1.4 (excluding the the signal drawn in red). At each iteration, the saliency S of the image I is obtained after it has been modified with respect to a chosen set of features. These features can be different from those used to compute saliency. The goal is to minimize the error between the saliency S and the target saliency T, given by $e = T - S$. In order to maintain the naturalness of the input
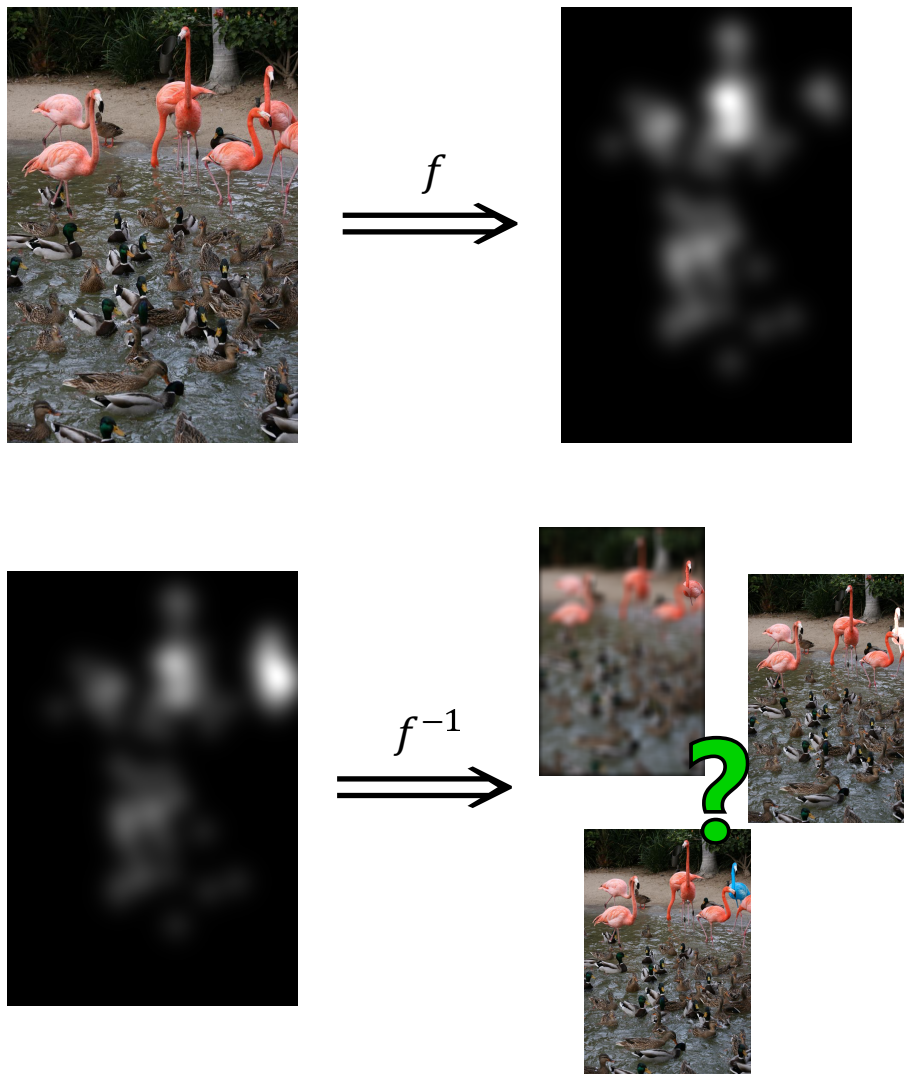
Figure 1.3: Image and corresponding saliency (top). Target saliency map and possible retargeted images (bottom).

image, a set of constraints can be introduced to prevent overmodification. The process ends when e becomes small enough, or when the modifications can no longer produce an appreciable change in saliency towards the intended goal.

An example of this simple approach can be found in the work of Wong and Low [40]. In their proposed model, the user divides an image into N segments and enumerates them in order of importance. A target importance value T_i is assigned to each segment, where $i = 1$ is the most important segment and $i = N$ is the least important. Modifications are made to the intensity, color saturation, and sharpness of the original image until the average saliency within each segment matches the target importance value for the corresponding segment.
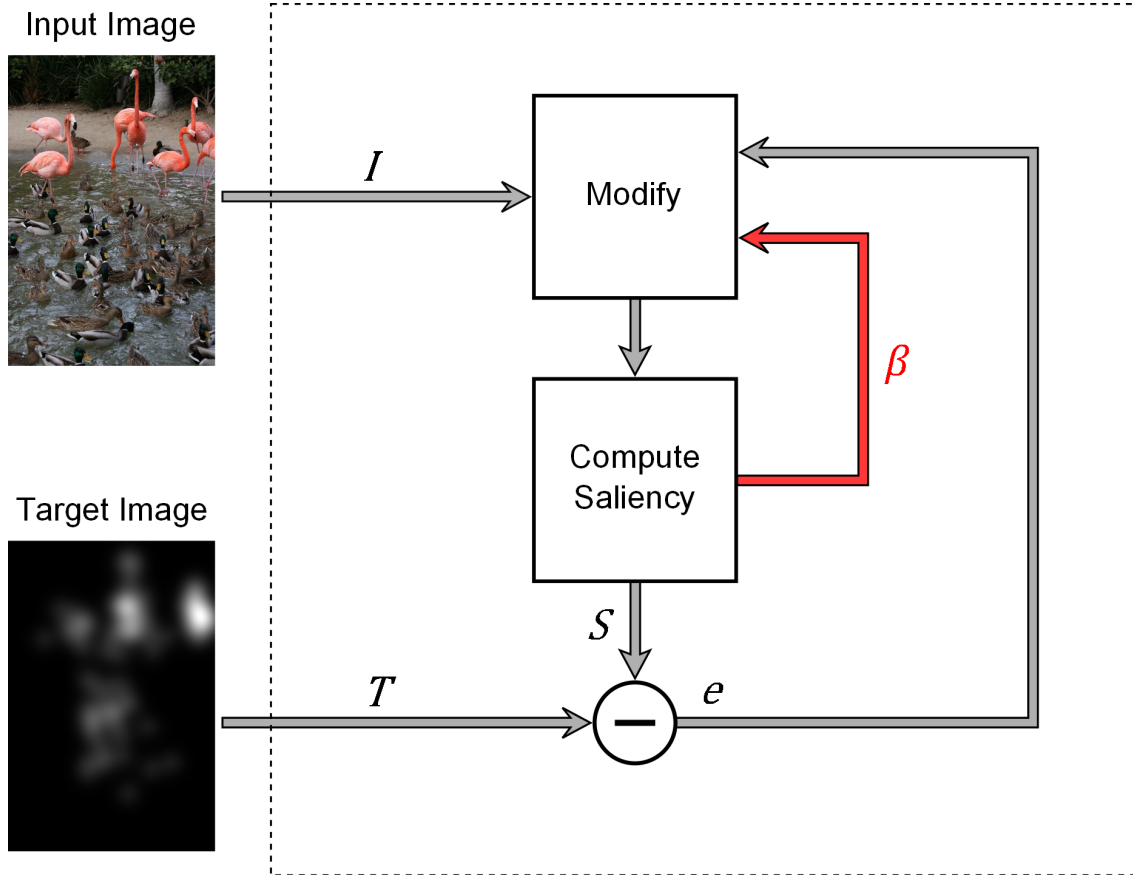
Figure 1.4: Attention Retargeting as an optimization procedure. The signal shown in red indicates that the model uses intermediate outputs of saliency computation (feature and conspicuity maps) to help determine the correct modification at each step.

Specifically, the saliency error is defined as

$$e = \sum_{i=1}^{N} |\mathcal{N}(T_i) - \mathcal{N}(S_i)|, \tag{1.1}$$

where S_i is the average saliency of the ith segment, and $\mathcal{N}(\cdot)$ is a normalization operator defined as $\mathcal{N}(X_i) = X_i / \sum_{j=1}^{N} X_j$. A set of bound constraints specifies the upper and lower bounds for the average intensity, color saturation, and sharpness of any segment. In addition, it is assumed that the intensity and color saturation should never be lowered on the most important segment, and that the more important the segment, the sharper it should be.

The main disadvantage of this approach is the limited insight into why the images are modified the way they are. This limitation stems from the use of saliency computation as a black-box function, which only lets us observe the effect of a modification after the change

to the image has been made. If $\mathcal{N}(T_i) - \mathcal{N}(S_i)$ is small for some i, no modifications are necessary in that region. If $\mathcal{N}(T_i) - \mathcal{N}(S_i)$ is large and positive then we have to increase saliency, and if $\mathcal{N}(T_i) - \mathcal{N}(S_i)$ is large and negative then we have to decrease saliency in order to match the target saliency. All we can do is modify the image where $|\mathcal{N}(T_i) - \mathcal{N}(S_i)|$ is large, recompute the saliency, and see whether the cost function is actually minimized. Since the modification may not necessarily alter saliency in the right direction, an excessive number of iterations may be needed.

1.2.2 Feedback from Saliency Computation

Rather than relying solely on the movement of the cost function to guide our modifications, we may take advantage of intermediate outputs from the saliency computation block if it contains feature channels that are pertinent to the modifications being made. For example, let us consider a saliency detector that extracts spatial frequency information from blocks of I to produce a feature map β_f. Center-surround operations are applied on β_f to produce the conspicuity map β_c, describing the saliency of I that originates from spatial frequency. The modification in this example is either blurring or sharpening (e.g., via Gaussian or Laplacian filters) applied to specific regions of I, which decrease or increase their spatial frequency, respectively, by varying degrees.

Consider the case where saliency must be reduced in a particular region. If β_c is small in this region, then the feature associated with it does not contribute to saliency and does not need to be modified. Conversely, if β_c is large then the feature in β_f must be altered in the opposite direction. In our example, if β_f indicates high spatial frequency in that region of I then we need to blur it; otherwise, we need to sharpen it. Similar logic is applied in the case where saliency must be increased in a particular region. A large β_c in this region indicates that it is already salient with respect to its corresponding feature, hence no modification is needed. If β_c is small in this region, then the corresponding region of I must be made salient with a modification that alters the feature in β_f in the opposite direction, as before. Thus, the information extracted during saliency computation $\beta = \{\beta_f, \beta_c\}$ can be used to guide the direction and possibly even the magnitude of the modifications at each step. This improved algorithm, shown by the addition of the red signal in Fig. 1.4, can be thought of as a steepest descent algorithm for attention retargeting.

Hagiwara *et al.* apply this methodology in their attention retargeting model [12]. They reverse engineer a simplified version of the well-known saliency model by Itti et al. [17], which decomposes an RGB image into a set of four color features $r = R - (G + B)/2$ for red, $g = G - (R+B)/2$ for green, $b = B - (R+G)/2$ for blue, and $y = (R+G)/2 - |R-G|/2 - B$ for yellow, and an additional intensity feature $v = (R + G + B)/3$. Conspicuity maps are then obtained for each feature and combined to compute saliency. In their model, Hagiwara *et al.* first select a ROI whose saliency is to be maximized. They determine the change in the image channels ΔR, ΔG, ΔB needed to increment saliency at each pixel as a function of the features and their corresponding conspicuity maps. Since the saliency of the ROI is to be maximized relative to the rest of the image, each pixel within the ROI is modified by $\Delta \alpha$ (calculated individually for each pixel), and pixels outside are modified by their respective $-\Delta \alpha$, where $\alpha \in \{R, G, B\}$.

1.2.3 Direct Mapping

The optimization procedure in Fig. 1.4 uses feedback from saliency computation to estimate how the image is to be modified in a series of steps. Mapping an additive or multiplicative change to the saliency map (or conspicuity map of a particular feature) directly onto the image domain is a highly non-trivial task. However, one of the earlier works on attention retargeting by Su *et al.* [33] demonstrates that this is indeed possible. Since there currently are no generalizations of this approach, we illustrate the concept of direct mapping by summarizing the methodology of Su *et al.*. Their method de-emphasizes distracting textures by decreasing the spatial variation in textured regions. In their approach, they compute the texture-based saliency of an image using steerable filter banks (Fig. 1.5) and then propagate changes in saliency backwards to the steerable coefficients. Since image decomposition through Gabor filters allows near perfect reconstruction, the new image is easily obtained from the modified steerable coefficients.

After the image is decomposed, the local frequency content for each subband s_n is computed. Since s_n is band-limited, full-wave rectification must be performed before local averaging with a Gaussian filter. The resulting average response s_{Ln} is called a "power map." The next step is to extract conspicuities within each subband, which is accomplished by applying a high-pass filter to each corresponding power map. The overall goal will be to use the coefficients in the texture conspicuity maps s_{Hn} to modify the steerable coefficients s_n
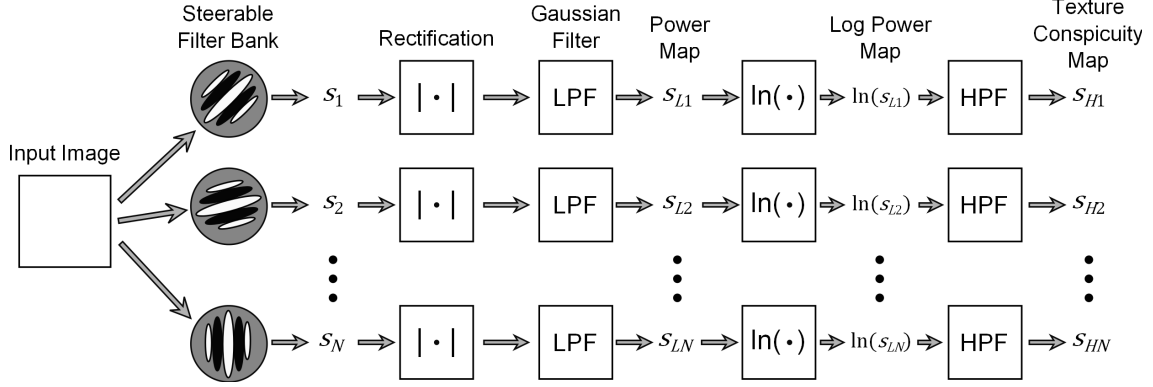
Figure 1.5: Computation of visual saliency from texture using steerable filter banks.

in a manner that produces a more uniform composition in highly-varying textured regions. This task can also be thought of as a modification to the local frequency content stored in the power map.

However, propagating the coefficients s_{Hn} backwards can be problematic due to the non-linear rectifying operation performed earlier. Note that all power map coefficients are positive and must remain as such after any modification. Therefore, the logarithm of the power map is taken prior to high-pass filtering and all subsequent modifications using the conspicuity map coefficients s_{Hn} are performed on the log power map to ensure that negative values cannot be produced on the power map. The desired modification is a reduction in texture variation, which corresponds to a removal of high frequency regions in the power map. This is accomplished simply by subtracting the conspicuity map coefficients s_{Hn} from the log power map coefficients $\ln(s_{Ln})$. Note that addition to the log power map coefficients corresponds to multiplication to the original steerable coefficients. Therefore, this modification can be propagated backwards directly onto s_n as a scaling operation with the coefficients s_{Hn}

$$s'_n = s_n e^{-k s_{Hn}}, \tag{1.2}$$

where k is some constant. The modified image can be easily reconstructed as the linear combination of the modified subbands s'_n. Unlike previous optimization-based approaches, this method manages to propagate additive changes directly from the saliency map back onto the image.

10

1.2.4 ROI-Based Retargeting

In many cases, it is only desirable to draw attention to a specific ROI, or perhaps a few ROIs. For example, the target saliency map in the bottom of Fig. 1.3 indicates that we want the flamingo in the top right of the image to draw more attention than the rest of the image. To this end, we may supply a simple binary mask of the desired ROI with the objective of making this region most salient relative to the rest of the image. The absence of the target saliency as an input highlights a big redundancy—the lack of need for full-scale saliency computation—in the approaches outlined in Section 1.2.1 for ROI-based attention retargeting. Saliency computation typically requires center-surround differences over the whole image. With our simplified goal of either increasing or decreasing saliency within a single region, it suffices to perform this operation on the ROI and its surroundings alone, rather than the entire image. Our models, presented in Chapters 2 and 3, illustrate the advantages of avoiding this inefficiency.

Aside from benefits to computational complexity, this simplification may also allow more freedom in the problem formulation, particularly in the objective function used in the optimization. For example, a recently proposed model by Nguyen *et al.* used a graph-based optimization for attention retargeting [29]. The input image is segmented into patches, and the set of patches i comprising the ROI undergo color transfer from a set of candidate patches x_i. The candidate patches are mined from a large image dataset and correspond to the same objects found in the ROI so that color transfer only occurs between similar objects in an effort to maintain naturalness. The ideal selection of candidate patches for color transfer to the ROI is found by minimizing

$$E(x) = \sum_{i \,\in\, \text{ROI}} E_d(x_i) + \lambda \sum_i \sum_{j \,\in\, N(i)} E_s(x_i, x_j). \tag{1.3}$$

The data cost E_d is designed to consider global center-surround differences, rather than local ones. It demands that the ideal candidate patch be highly dissimilar from the entire image content outside of the ROI. The smoothness energy E_s is designed to punish dissimilarity between neighboring patches that are of a similar color, e.g., belonging to the same object, and encourage it otherwise. In contrast to the data cost E_d, this considers local center-surround differences to further enhance saliency. At the same time, it ensures that different objects in the ROI remain individually consistent in appearance.

1.3 Organization

This thesis is organized as follows. In Chapter 2, we present our method for attention retargeting by directly modifying the orientation of a ROI in an image. We also evaluate our method using eye-tracking to confirm its effectiveness in drawing attention. Our second attention retargeting method, which operates on the color of a ROI in an image, is presented in Chapter 3, with a similar eye-tracking evaluation. An additional subjective experiment is conducted to verify the naturalness of the results. Chapter 4 discusses the plausibility of subliminal attention guiding in natural images. We describe our experiments to determine whether subliminal flicker in natural images is capable grabbing a viewer's attention and present some preliminary results. Conclusions and suggestions for future work are presented in Chapter 5.

Chapter 2

Attention Retargeting by Manipulating Orientation

This chapter describes the attention retargeting model we proposed in [23], which estimates the relative changes in saliency of a ROI if it were to be rotated. This is accomplished by summarizing the orientation content of a selected ROI and its local surroundings using the procedure described in Section 2.2. The section that follows shows how a rotation of the ROI can be concisely represented so that center-surround differences can be quickly computed for all possible rotations. Finally, we apply our method to a set of natural images and verify its effectiveness in guiding attention through eye-tracking.

2.1 Iterative Approach

Suppose we apply the iterative approach described in Section 1.2.1 with the goal of modifying the saliency of an image by manipulating its orientation. This procedure is summarized below in Algorithm 1. The saliency computation in step 7) would typically include an orientation feature channel, which for example extracts orientation information using Gabor filters. These feature maps and corresponding conspicuity maps can be used as described in Section 1.2.2 to determine more accurately the direction of rotations in step 4) of Algorithm 1.

This procedure is unsatisfactory for several reasons. First, the concept of rotating every individual piece of an image in order guide attention to a particular region (or perhaps a few) is quite silly, and would most likely produce results suitable only for an abstract

---

Algorithm 1

1: Segment the original image into N disjoint segments (e.g., using superpixel segmentation).

2: Initialize the rotation angle of the ith segment $\phi_i = 0°$, where $i = 1, ...N$.

3: For each segment i

4:     change the rotation angle $\phi_i = \phi_i \pm \Delta\phi$.

5:     rotate the ith segment in the original image by ϕ_i degrees.

6: Inpaint the gaps produced by the rotations.

7: Compute the saliency S of the resulting image.

8: If the error between saliency S and target saliency T is small enough, stop; otherwise, go to step 3).

---

art project. The noise introduced by having to inpaint gaps left by rotating N different segments further compounds this issue. Second, the iterative nature and the need to run inpainting in step 6) and saliency computation in step 7) may lead to high computational complexity. This will likely be the case even if we restrict our rotations to only a small subset of segments, leaving the rest unmodified. And finally, the procedure (if successful) does not provide any direct insight as to why the chosen rotation angles draw our attention to the desired areas of the image.

We propose a method of computing relative ROI saliency directly for any given rotation angle. Since we only wish to draw attention to a particular region, we apply the methodology in Section 1.2.4. This enables us to efficiently predict the rotation angle at which the relative saliency is maximized without any iterative procedure—a previously unexplored concept. In addition, our procedure makes it clear why saliency is maximized as intended.

2.2   Edge Distributions

Our first task is to obtain distributions that specify the occurrence of edges over a range of orientations in the ROI and its surroundings. A well-established method of detecting collinearity among feature points in an image, referred to here as the standard Hough Transform, was developed decades ago [14, 7]. A mapping of point coordinates onto a discrete 2D parameter space (the so-called Hough space) is performed based on the polar representation of a line:

$$\rho = x \cos\theta + y \sin\theta. \tag{2.1}$$

14

Given the coordinates, x and y, of a pixel in the image, the parameter ρ is solved using (2.1) for all $\theta \in [0, 180°)$ to yield a sinusoidal curve in the Hough space. The points at which these curves intersect indicate the angle-radius parameters of lines that best fit the set of input pixels in the image. In practice, the Hough space is a 2-D histogram whose bins are incremented at each point of the sinusoidal curve obtained from (2.1) over a set of angles $\theta \in [0, 180°)$.

In our method, we wish to treat this 2D histogram as a joint probability density function (pdf) $p(\theta, \rho)$ of the line parameters so that we can obtain the marginal density $p(\theta)$, which describes the likelihood that an edge in the region is oriented at an angle $\theta \in [0, 180°)$. Unfortunately, the standard Hough transform is not well-suited for this operation. Since each point (x, y) contributes a single vote in the histogram for all values of θ, marginalizing the histogram yields a uniform distribution with a height equivalent to the number of input points considered.

To overcome this problem, we utilize a statistical Hough transform [5], which uses kernel densities to estimate the edge content in the Hough space as a continuous pdf. At each pixel i with coordinates (x_i, y_i) the orientation θ_i of the pixel can be estimated as

$$\theta_i(x_i, y_i) = \arctan\left(\frac{I_y(x_i, y_i)}{I_x(x_i, y_i)}\right), \tag{2.2}$$

where I_x and I_y are the two components of the spatial gradient of the image intensity $I(x, y)$. Intuitively, if the pixel i lies on a strong edge, we can be more confident in our estimate of θ_i, hence the uncertainty of our estimate is inversely proportional to the gradient magnitude

$$\sigma_{\theta_i}(x_i, y_i) = \frac{1}{\sqrt{I_x^2(x_i, y_i) + I_y^2(x_i, y_i)}}. \tag{2.3}$$

Using the set of observations $\{(\theta_i, x_i, y_i)\}_{i=1\ldots M}$ we model the edge content in the Hough space as proposed in [5]. The joint pdf of the line parameters (θ, ρ) and the pixel coordinates (x, y) is expressed as

$$p(\theta, \rho, x, y) = p(\rho|\theta, x, y)p(\theta, x, y). \tag{2.4}$$

However, ρ is deterministic through (2.1) when x, y, and θ are known, hence [5] models the conditional pdf using the Dirac delta function as follows:

$$p(\rho|\theta, x, y) = \delta(\rho - x\cos\theta - y\sin\theta). \tag{2.5}$$

15

The remaining term in (2.4) is estimated given the set of observations $\{(\theta_i, x_i, y_i)\}_{i=1\ldots M}$ by using kernels [5]:

$$p(\theta, x, y) = \sum_{i=1}^{M} \frac{1}{\sigma_{x_i}} K_x \left(\frac{x - x_i}{\sigma_{x_i}} \right) \frac{1}{\sigma_{y_i}} K_y \left(\frac{y - y_i}{\sigma_{y_i}} \right) \frac{1}{\sigma_{\theta_i}} K_\theta \left(\frac{\theta - \theta_i}{\sigma_{\theta_i}} \right), \qquad (2.6)$$

where p_i is the prior on the observation (θ_i, x_i, y_i). In our implementation, we set K_x, K_y, and K_θ to be Gaussian kernels. Due to the single pixel precision of the image coordinate plane, we set the widths of the spatial kernels $\sigma_{x_i} = \sigma_{y_i} = 1$, for all i. We also consider all observations within the ROI to be equiprobable, i.e., $p_i = 1/M$. The analysis of the region surrounding the ROI needs to emphasize local data, since this is what governs low-level saliency. We model this by convolving a Gaussian function with a full width at half maximum of 5° of visual angle along the ROI's border. The priors are then formed by normalizing the resulting grayscale map such that $\sum_{i=1}^{M} p_i = 1$.

After plugging (2.5) and (2.6) back into (2.4) and integrating with respect to variables x and y, an estimate of the Hough transform is obtained

$$p(\theta, \rho) = \sum_{i=1}^{M} \frac{1}{\sigma_{\theta_i}} K_\theta \left(\frac{\theta - \theta_i}{\sigma_{\theta_i}} \right) R_i(\theta, \rho) p_i, \qquad (2.7)$$

where $R_i(\theta, \rho)$ is the Radon transform of the spatial kernels

$$R_i(\theta, \rho) = \int \int \delta(\rho - x \cos \theta - y \sin \theta) \frac{1}{\sigma_{x_i}} K_x \left(\frac{x - x_i}{\sigma_{x_i}} \right) \frac{1}{\sigma_{y_i}} K_y \left(\frac{y - y_i}{\sigma_{y_i}} \right) dx \, dy. \qquad (2.8)$$

The pdf in (2.7) is integrated with respect to ρ to obtain our "edge distribution,"

$$p(\theta) = \int p(\theta, \rho) d\rho. \qquad (2.9)$$

2.3   Relative ROI Saliency Prediction

The procedure in Section 2.2 is performed independently on the observations within the ROI and outside the ROI to obtain their respective edge distributions, $p_R(\theta)$ and $p_S(\theta)$. The example in Fig. 2.1(b) illustrates the edge distribution for the ROI (solid red plot) and its surroundings (solid blue plot) in the image shown in Fig. 2.1(a). The large peaks in $p_R(\theta)$, located around 0°/179.5° and 90°, indicate that the tile in the selected region is predominantly composed of horizontal and vertical edges. The plot of $p_S(\theta)$ indicates a similar composition of edges in its surroundings. If the ROI were rotated in a counter-clockwise direction by 45°, for example, the tile's edges would then be oriented primarily

16

(a)

(b)

(c)
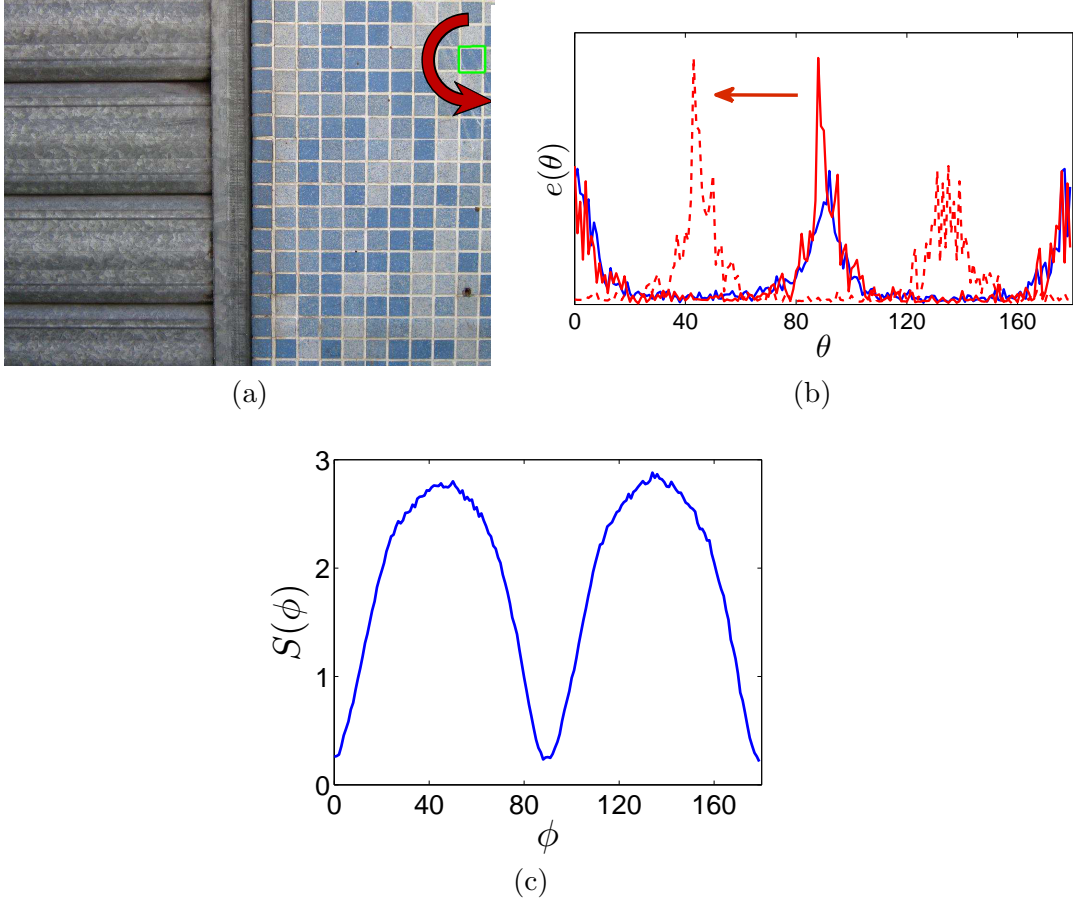
Figure 2.1: Illustration of the method used to predict relative saliency of the ROI in (a): (b) shows the edge distributions of the ROI (red) and surround (blue); (c) shows the KLD $S(\phi)$ between ROI and surround edge distributions as a function of rotation angle ϕ.

around $45°$ and $135°$. Hence, a counter-clockwise rotation of the ROI can be represented as a leftward circular shift of $p_R(\theta)$, as illustrated in Fig. 2.1(b) by the dashed red plot.

The region's relative saliency is predicted by the dissimilarity between the two edge distributions, which is measured as their symmetric Kullback-Leibler (KL) divergence. Specifically, the predicted relative ROI saliency for rotation angle ϕ is obtained as

$$S(\phi) = \frac{1}{2}D\left(p_S(\theta)||p_R(\theta+\phi)\right) + \frac{1}{2}D\left(p_R(\theta+\phi)||p_S(\theta)\right), \qquad (2.10)$$

where $D(\cdot||\cdot)$ is the KL divergence [22]. An example of the predicted saliency of the ROI in Fig. 2.1(a) as it is rotated counterclockwise is shown in Fig. 2.1(c). As expected, the saliency of the region peaks when its edges are most dissimilar to the edges in its surroundings—at rotations of about $45°$ and $135°$. Furthermore, its saliency is predicted to decrease back to

17

its original level as the rotation approaches 90°, at which point its edges would be oriented similarly to the edges in its surroundings.

This demonstrates what can be accomplished by avoiding full-scale saliency computation, as mentioned in Section 1.2.4. We can rapidly compute the center-surround difference of the ROI due to the concise representation of this modification as a circular shift of a 1-D distribution. The result, shown in Fig. 2.1(c), is a complete inverse mapping of relative ROI saliency to rotation angle (note that a one-to-one mapping between saliency and images still does not exist, despite only a single feature being modified). The approach taken here chooses features in a manner that allows a simpler representation of the modification being made, and utilizes a center-surround operation that avoids redundancy by only focusing on the ROI. These two aspects grant the ability to rapidly predict the outcome of all possible modifications of the particular feature, in this case orientation.

2.4   Evaluation and Results

We applied our method on 40 natural images from a dataset collected by Judd et al. [20], which includes eye-tracking data from 15 viewers. In each image, we used this data to help select an uninteresting region located away from areas that viewers generally fixated upon. To speed up computations, only edge pixels, identified by Canny edge detection [2], were considered in our analysis of the area outside of the ROI. Once the predicted relative ROI saliency was obtained, we rotated the ROI in each image to the angle at which the relative saliency was maximized and filled the missing background regions using the inpainting method from [4].

Eye-tracker data was collected from a total of 24 nonexpert participants using a head-mounted Locarna "Pt-Mini" eye-tracker. Participants were divided into two groups of 12. One group viewed the modified image set, while the other viewed the unmodified set. In either test, each image was displayed for 5 seconds in fixed order with a 3-second pause in between. During the 3-second pause, a small crosshair was shown on the center of the display and participants were asked to fixate on it. This was done to ensure that the current image did not affect the gaze data of the image that followed. Participants were seated 80 cm from a 19″ display with a native resolution of 1280×1024 in a room with an ambient
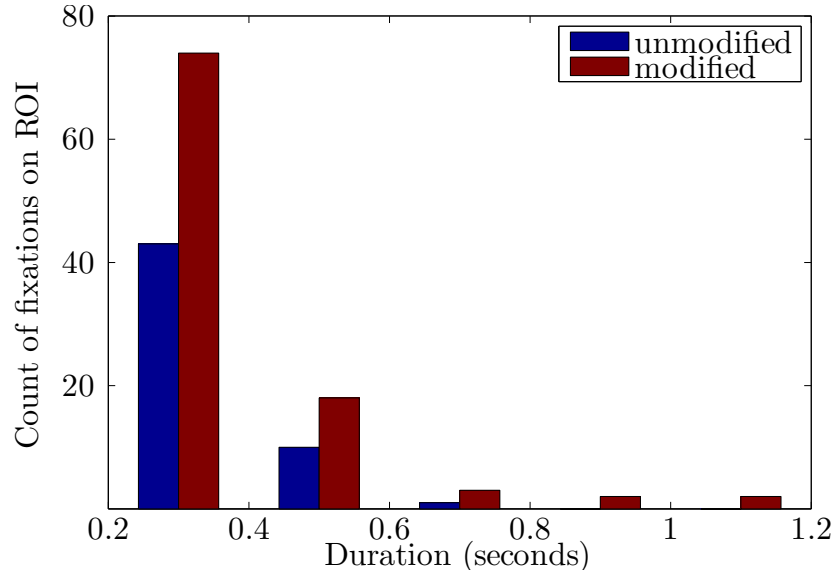
Figure 2.2: Histograms of the duration of fixations near the ROI.

light of 180 lx. The resolution of the display was chosen so that all of the images could be viewed at their native resolution.

Gaze data can be divided into two categories: fixations, in which the viewer's gaze has stabilized on a single location, and saccades, in which the viewer's gaze rapidly moves towards a new location. We define a fixation as a set of consecutive data points that are within a proximity of 1° of visual angle for a minimum of 0.2 seconds [6]. A fixation was determined to be close to the ROI if it was located within a certain radius from the ROI's centroid. This radius was chosen to be 5% of the image's diagonal, which on average corresponded to 63 pixels or 1.3° of viewing angle. Since fixations can vary in length, it is inadequate to simply count them, given that more value should be placed towards longer fixations. Histograms of the duration of fixations close to the ROI for either group are shown in Fig. 2.2. The histogram corresponding to the control group (participants looking at unmodified images) is shown in blue, while that for the test group (participants looking at modified images) is shown in red. An increase in the overall number of fixations by a factor of 1.83 is observed in the test group relative to the control (99 fixations out of a total of 4024 across all images for the test group and 54 out of 4053 for the control). We also observe that viewers are more likely to fixate on the ROI of the modified images for longer periods of time.

In addition to drawing attention directly onto the ROI, our goal is to guide the viewer's attention towards the ROI's general vicinity. An increase in the relative saliency of a ROI can be considered successful if we observe a shift in the gaze pattern towards the ROI. Some examples of these gaze shifts are illustrated using fixation heatmaps in Fig. 2.3. A video showing several examples of gaze data visualizations on modified images in our experiment has been made available at [41].

2.5   Conclusion

Orientation, though one of the most fundamental visual features that drive early attention, has never been previously considered in attention retargeting. In this chapter, we propose a method to predict the amount of attention a ROI is expected to draw after it has been rotated. Existing iterative approaches are inadequate for our goals since they are unintuitive and likely carry large computational costs, even when a single ROI is considered. In our novel approach, we analyze the orientation content of a region and its local surroundings and measure the difference between the two as the region's orientation is modified. We select uninteresting regions in a set of natural images and rotate each one at the
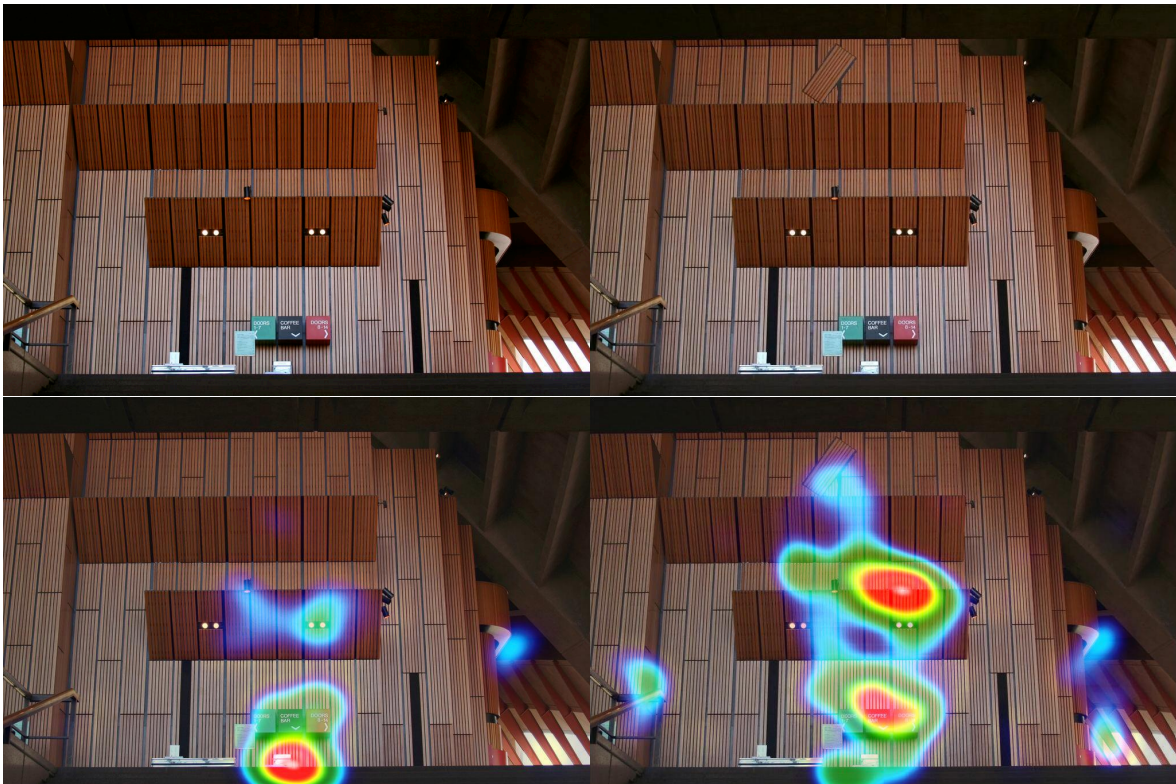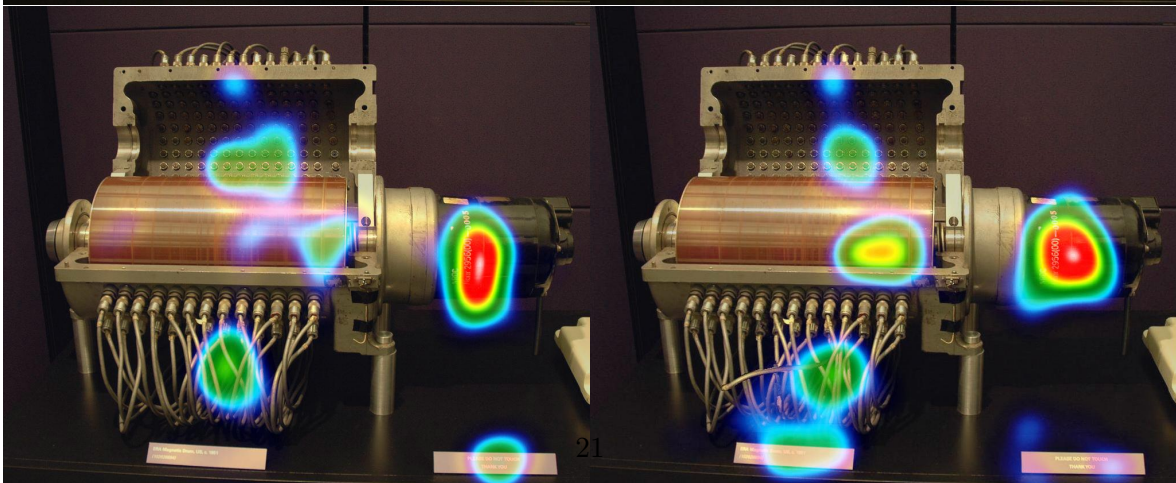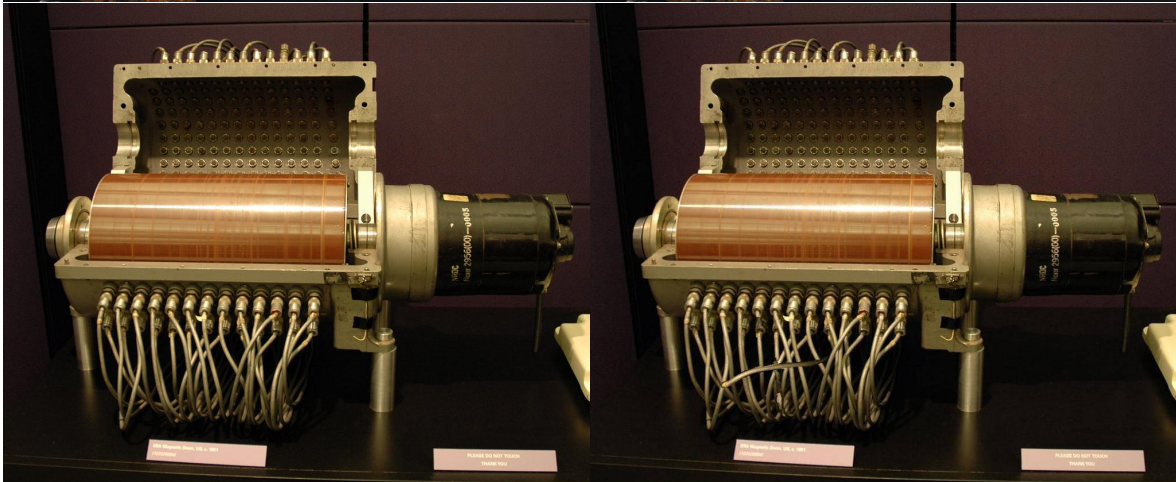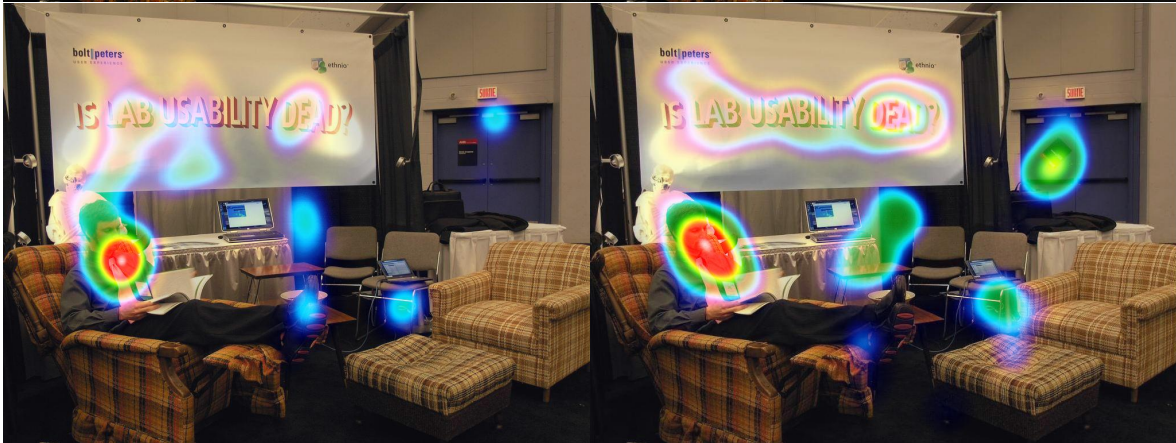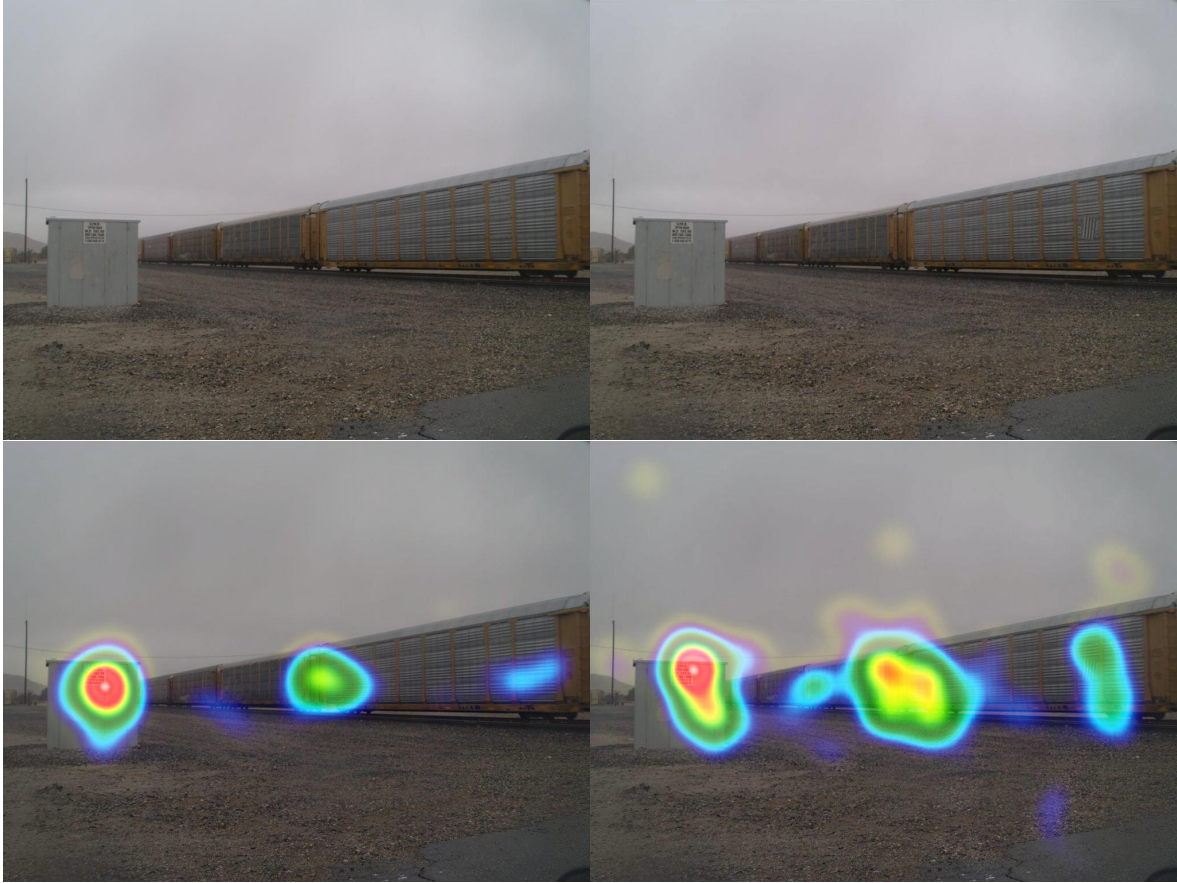
Figure 2.3: Examples of original and modified images used in our experiments with heatmap visualization of fixations. Each 2-by-2 array that corresponds to an image is organized as follows: original image (top-left); modified image (top-right); heatmap of fixations for original image (bottom-left); heatmap of fixations for modified image (bottom-right).

angle where the region's relative saliency is predicted to be maximized, filling in the missing gaps with inpainting. The eye-movements of participants who viewed the modified images are compared against a control group who viewed the original images. We demonstrate an overall increase in the number of fixations near the modified regions as well as an increase in their duration. Furthermore, we observe shifts in gaze towards the general vicinity of the modified region, which suggests that the region's saliency was successfully augmented.

Chapter 3

Attention Retargeting by Color Manipulation

The methodology from the previous chapter is extended to maximize the saliency of a selected ROI by modifying its color, with an additional focus on maintaining the naturalness of the original image. The first section provides motivation and discusses the underlying concepts. The rest of the chapter is organized similar to the previous, with an added subjective experiment to gauge the naturalness of the images produced by the proposed method. The material in this chapter has been presented in [24].
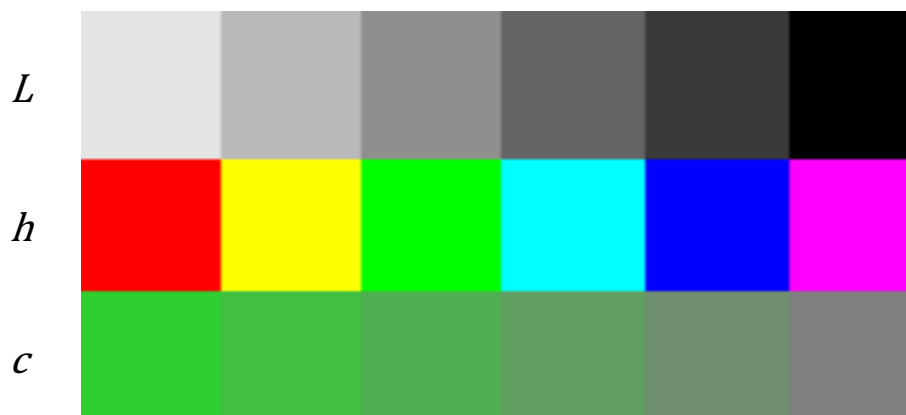
Figure 3.1: Illustration of the relevant attributes of color: Lightness (top), hue (middle), and chromaticity (bottom).

<div align="center">(a)   (b)   (c)</div>
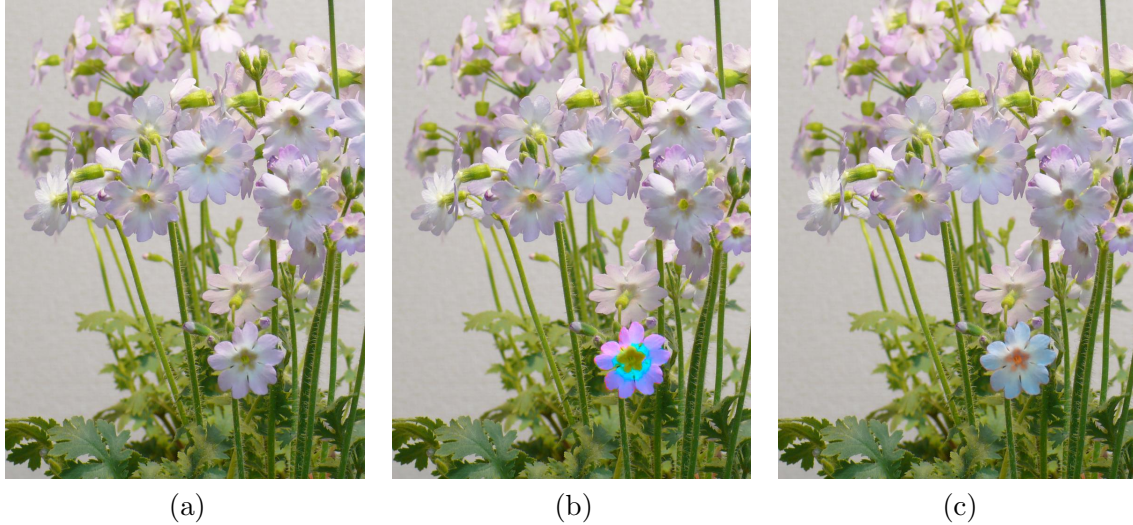
Figure 3.2: Chromaticity-based (b) and hue-based (our result) (c) saliency enhancement for the original image (a).

3.1 Basic Principles

To begin, we explain the various attributes of color relevant to this chapter. Lightness (or intensity) is the overall amount of light that is perceived to be emitted from a stimulus. Hue is the extent to which a color resembles red, blue, yellow, or green, or a combination of these. Chromaticity (or saturation) refers to the amount of white light that dilutes a hue.

Generally, when one thinks of color, they think of it in terms of hue as described above. In this work, we will only focus on modifying the hue component of color while leaving other attributes unaltered. Although lightness and chromaticity modifications can significantly alter saliency, this usually comes at a hefty price—a destructive effect on the image's perceived naturalness. An example of this is shown in Fig. 3.2(b), where the chromaticity of the ROI has been raised to enhance its saliency. The modification to the ROI in Fig. 3.2(b) is likely to draw attention, perhaps even more so than that of our result in Fig. 3.2(c). However, there is a sense that the flower in Fig. 3.2(b) disobeys the overall structure of lighting and color in the image and looks unnatural, especially compared to the corresponding flower in Fig. 3.2(c). These issues with lightness and chromaticity are what befall the methods in [40] and [12].

Of course, restricting modifications to hue only does limit our approach in certain cases. However, manipulating lightness and chromaticity in a responsible manner that maintains naturalness is a burden that raises complexity, which runs counter to our goal of developing

a fast, efficient method of saliency enhancement. Finally, we aim to provide insight as to why the modifications to color increase saliency—a task best accomplished by focusing on hue, the attribute that best characterizes what we perceive as color.

To accomplish our goals, we utilize CIE *Lch* color space, the polar representation of CIE *L*a*b** [32]. The three components of the *Lch* color space relate to *L*a*b** as:

$$L = L^*, \quad c = \sqrt{(a^*)^2 + (b^*)^2}, \quad h = \arctan \frac{b^*}{a^*}. \tag{3.1}$$

In addition to decoupling lightness, chromaticity, and hue, this color space has the advantage of being perceptually uniform. This means that the Euclidean distance between two points within the space directly corresponds to the perceived difference between the colors at those points [18].

Suppose that we are given a homogeneously-colored image with hue h_0 and a selected ROI. Since saliency at the lowest level is governed by local conspicuities in principal visual features, it then follows that the adjustment that maximizes saliency in the ROI is the one that makes it most dissimilar to its surroundings. By fixing L and c in (3.1), this adjustment is a 180° rotation to the hue, which results in the largest possible displacement in the color space while keeping L and c constant (see Fig. 3.3). Any out-of-gamut colors that result from this modification need to be clamped so that they can be represented in sRGB color space.

Of course, natural images are usually not composed of a single hue. Both ROI and its surround are likely to contain many different hues. A convenient way to represent the hue content of ROI and/or its surround is through the hue histogram or hue distribution. If there are many pixels at a certain hue, there will be a peak at that hue value in the hue distribution. Given the hue distributions in the ROI and the surround, the task is to obtain ROI hue distribution that is most dissimilar to the hue distribution in the surround. This problem is analyzed in the following subsections.

3.2 Hue Distributions

We model the hue information within the ROI and its surroundings by the corresponding hue distributions while taking into account the uncertainty inherent in the representation of hue. Note that as chromaticity c decreases towards 0 in equation (3.1), meaning that
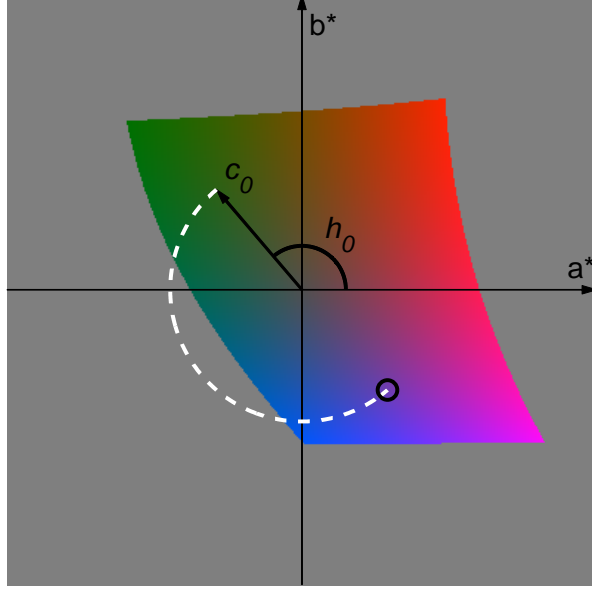
Figure 3.3: Optimal hue adjustment to maximize the saliency of a ROI in a homogeneously-colored image.

$a^*, b^* \to 0$, the value of hue h becomes uncertain. Physically, this means that a shade of gray is hard to categorize as red, blue, green, or yellow.

To obtain a measure of this uncertainty, we follow the route taken by [10]. Consider a function of several measurements $u = q(\hat{u}_1, \hat{u}_2, ..., \hat{u}_N)$. For independent measurements with corresponding standard deviations $\sigma_{\hat{u}_1}, ..., \sigma_{\hat{u}_N}$, the uncertainty in u is given by [34]

$$\sigma_u = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial q}{\partial \hat{u}_i} \sigma_{\hat{u}_i} \right)^2}. \tag{3.2}$$

Applying this to h in (3.1) with $\sigma_{a^*}^2 = \sigma_{b^*}^2 = \alpha$, we obtain

$$\sigma_h = \frac{\alpha}{c}. \tag{3.3}$$

Since we are only concerned with the uncertainty originating from the structure of hue (i.e., $\sigma_h \to \infty$ as $c \to 0$), we set α equal to the largest value of chromaticity of all pixels present in the image. In addition to limiting σ_h within the range $[1, \infty]$, which is a nice property for kernel density estimation, this makes σ_h a relative measure of uncertainty for each image. This is reasonable since we would expect relevant observations in vibrantly colorful images to come from pixels with high chromaticity, whereas more lenience would be given in the case of bland, faded images.
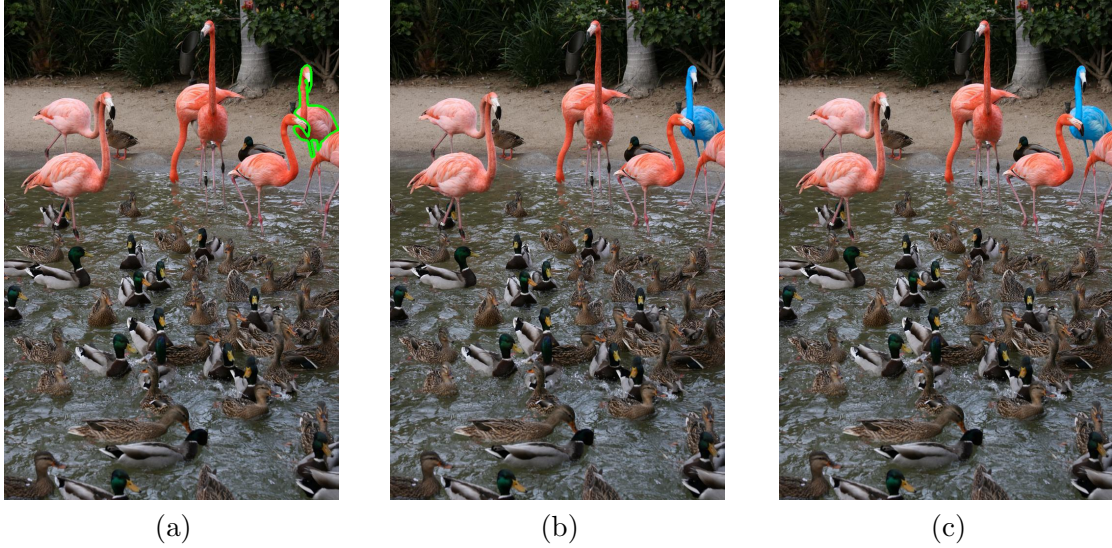
Figure 3.4: Example image showing the ROI outlined in green (a), and the hue-adjusted ROI using Gaussian (b) and Dirac (c) kernel density estimates.

For $h \in [0, 360°)$ degrees, the hue distribution can be constructed using kernel density estimation

$$p(h) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\sigma_{h_i}} K \left(\frac{h - h_i}{\sigma_{h_i}} \right) p_i, \tag{3.4}$$

where M is the total number of pixels under consideration, h_i is the hue observed at pixel i with corresponding uncertainty σ_{h_i}, and p_i is the prior. We considered Gaussian ($\mathcal{N}(h_i, \sigma_{h_i}^2)$) and Dirac ($K(h) = \delta(h - h_i)/\sigma_{h_i}$) kernels in this work. Using the Dirac kernel turns the kernel density estimation (3.4) into a simple task of computing a weighted histogram, which is significantly faster, especially since one has to consider the wrap-around property of hue ($h + 360° = h$) when computing $p(h)$ with Gaussian kernels. When analyzing the ROI, we consider each observation h_i to be equiprobable, i.e., $p_i = 1/M$. The priors of the surrounding region are modeled just as in Section 2.2. A Gaussian function with a full width at half maximum of 6° of visual angle is convolved around the ROI's border. The resulting grayscale map is normalized such that $\sum_{i=1}^{M} p_i = 1$ to give the values of the priors.

3.3 Optimal Hue Adjustment

Let the obtained hue distributions of ROI and surround be denoted $p_R(h)$ and $p_S(h)$, respectively. Consider the image in Fig. 3.4(a), where the ROI is outlined in green. Fig. 3.5 shows the distribution of hues for the ROI, indicated in red, and surroundings, indicated
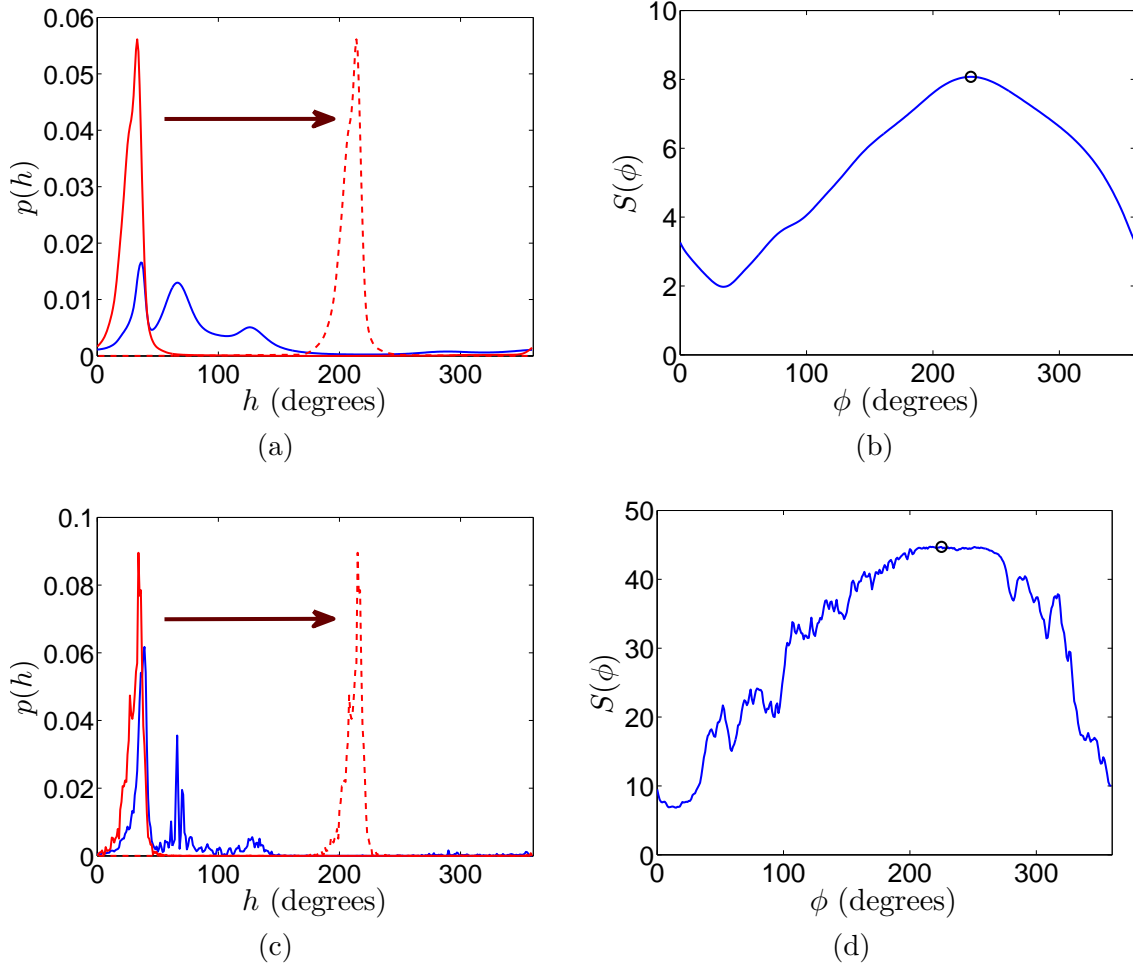
Figure 3.5: Illustration of hue adjustment using Gaussian (top) and Dirac (bottom) kernel density estimates: (a) and (c) show hue densities of the ROI (red) and surround (blue); (b) and (d) show KLD $S(\phi)$ between ROI and surround hue densities as a function of hue shift ϕ.

in blue, obtained using Gaussian and Dirac kernel density estimates (Fig. 3.5(a) and Fig. 3.5(c), respectively). Using this representation, a clockwise rotation ϕ of the hue in the CIE Lch space for all observations within the ROI corresponds to a rightward circular shift of $p_R(h)$ as illustrated in Fig. 3.5(a) and (c) by the dashed red lines. Following the discussion in Section 3.1, the adjustment that maximizes the saliency of the ROI relative to its surroundings is the one that makes $p_R(h)$ most dissimilar to $p_S(h)$. We measure the dissimilarity between distributions as their KL divergence:

$$S(\phi) = D\left(p_S(h) \| p_R(h + \phi)\right). \tag{3.5}$$

28

The plots in Fig. 3.5(b) and (d) are obtained by computing (3.5) with the corresponding distributions in Fig. 3.5(a) and (c) for all adjustments ϕ in 1° steps. Since Gaussian kernels tend to produce smooth, continuous distributions, we generally find that the resulting $S(\phi)$ has a clearly-defined maximum. Thus, the optimal adjustment in this case is simply the one that maximizes $S(\phi)$, i.e., $\phi_{\text{opt}} = \arg\max S(\phi)$.

When using Dirac kernels, it is common to obtain several different values of ϕ that maximize $S(\phi)$ due to flat peaks. An example is shown in Fig. 3.5(d), where it appears that any ϕ in the interval $[195°, 270°]$ degrees is suitable. A reasonable approach would be to take the median of this set as the solution. This procedure is summarized in Algorithm 2. The optimal hue adjustment obtained with this procedure using $k = 0.9$ on the $S(\phi)$ of Fig. 3.5(d) is marked with a circle.

Algorithm 2

 1: Select threshold $k \leq 1$.
 2: Find the set $\Phi = \{\phi : S(\phi) \geq k \cdot \max S(\phi)\}$.
 3: For each connected interval Φ_i in Φ (taking into account the periodicity of hue)
 4:     calculate the the length $||\Phi_i||_1$.
 5:     calculate $\phi_{\text{opt},i} = \text{median}(\Phi_i)$.
 6: $\phi_{\text{opt}} = \phi_{\text{opt},i^*}$, where $i^* = \arg\max ||\Phi_i||_1$.

Once ϕ_{opt} is obtained, all observations within the ROI are adjusted as $h'_i = h_i + \phi_{\text{opt}}$. All out-of-gamut pixels are clamped when converting the image back to sRGB color space. Note that the result of hue adjustment with Dirac kernels and Algorithm 2, shown in Fig. 3.4(c), very closely approximates the result obtained using Gaussian kernels in Fig. 3.4(b), with the advantage of being much faster.

3.4 Evaluation and Results

We applied our method to 34 images taken from the datasets by Judd *et al.* [20] and Nguyen *et al.* [29] (which were resized by a factor of 1.6 using bicubic interpolation). The Dirac kernel-based procedure was utilized with $k = 0.9$ to modify the ROIs. A MATLAB implementation of our method is available at www.sfu.ca/~ibajic/software.html.
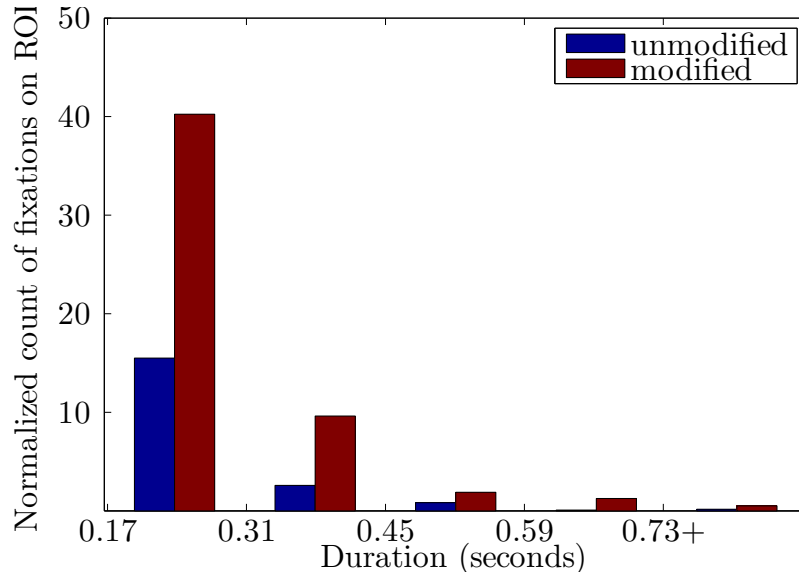
Figure 3.6: Histograms of the mean duration of fixations within the ROI per viewer.

3.4.1 Eye-Tracking Tests

Eye-tracking data was collected from a total of 20 nonexpert participants using a head-mounted Locarna "Pt-Mini" eye-tracker. The control group, consisting of 12 participants, viewed the unmodified images, while the remaining participants (test group) viewed the modified images. Participants placed their head on a chin rest at a distance of 70 cm from a 19″ display with a native resolution of 1280×1024 in a room with an ambient light of 190 lx. Apart from the initial resizing of the images from Nguyen *et al.* [29], all images were shown at their native resolution. Each image was displayed for 4 seconds in fixed order between two 1.5-second pauses.

A set of consecutive gaze data points are classified as a fixation if they lie within a proximity of 1° of visual angle from each other for a minimum of 0.167 seconds [6]. To account for possible eye-tracking errors, we dilate the binary mask of the ROI for each image using a circular disk with a radius of 0.5° of visual angle and consider a fixation to be on the ROI if it is located within this dilated mask. Histograms of the mean duration of fixations on the ROI per viewer for either group are shown in Fig. 3.6, with blue corresponding to the control group and red corresponding to the test group.

A histogram count of the duration of fixations on the ROI for each group was computed and subsequently normalized by the number of participants in each group. The histogram shows an average of 53 and 19 fixations on the ROI per viewer (out of an average total
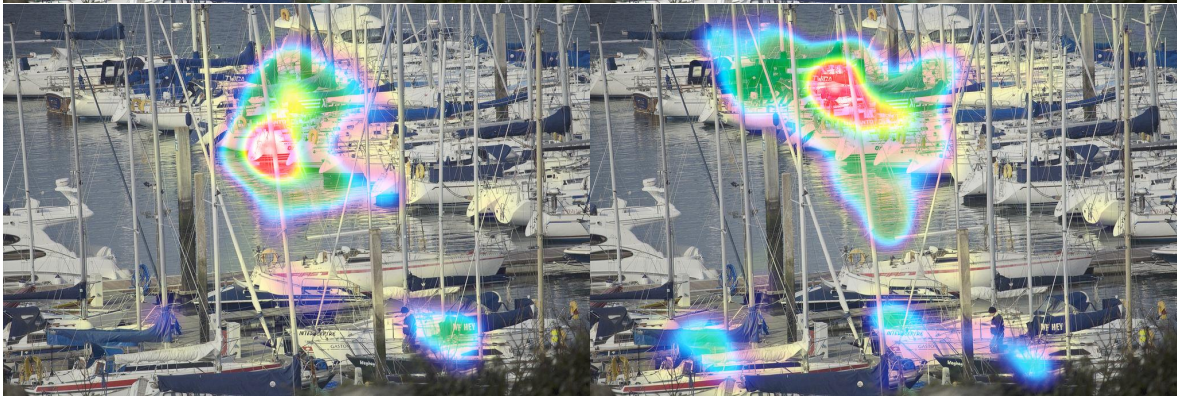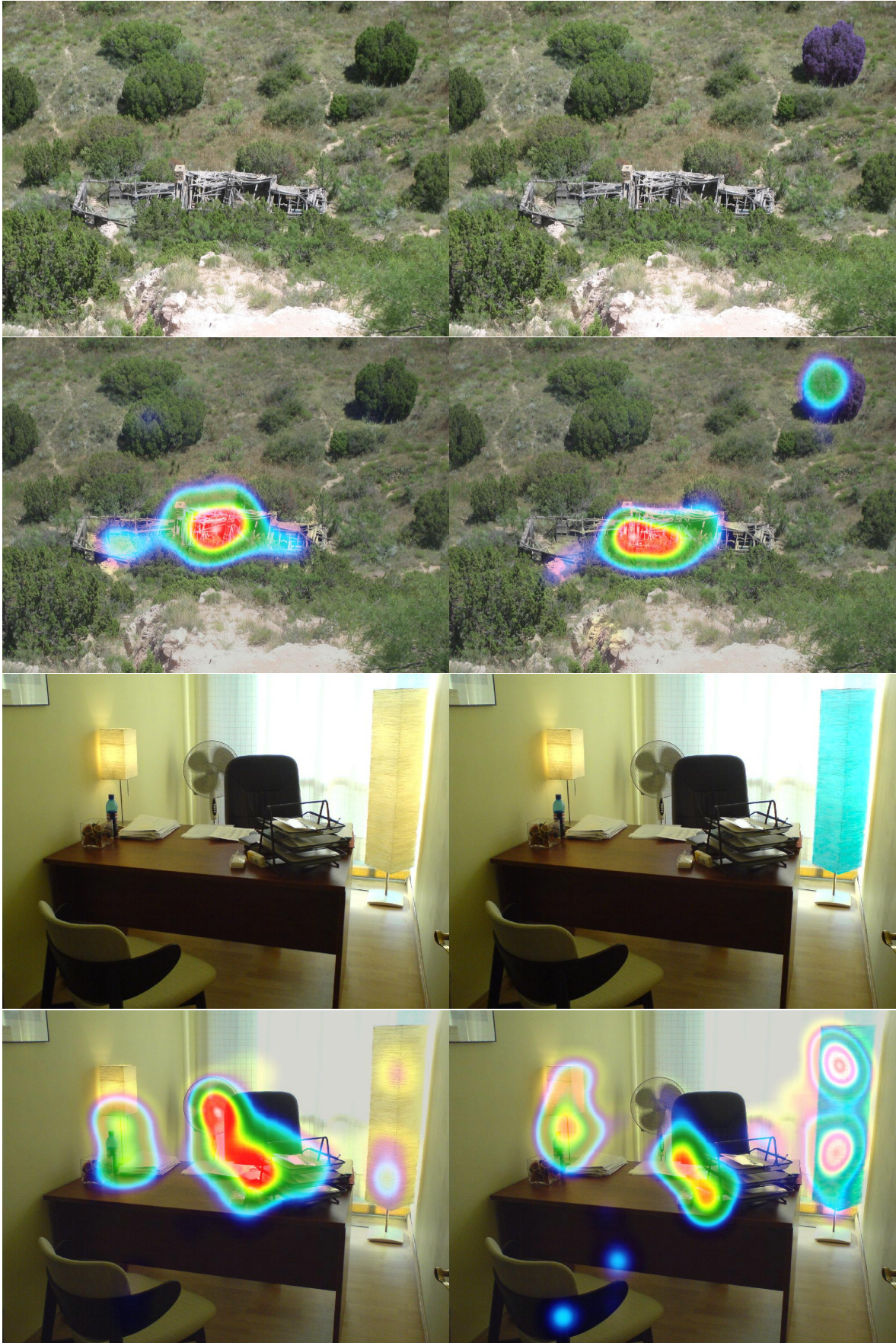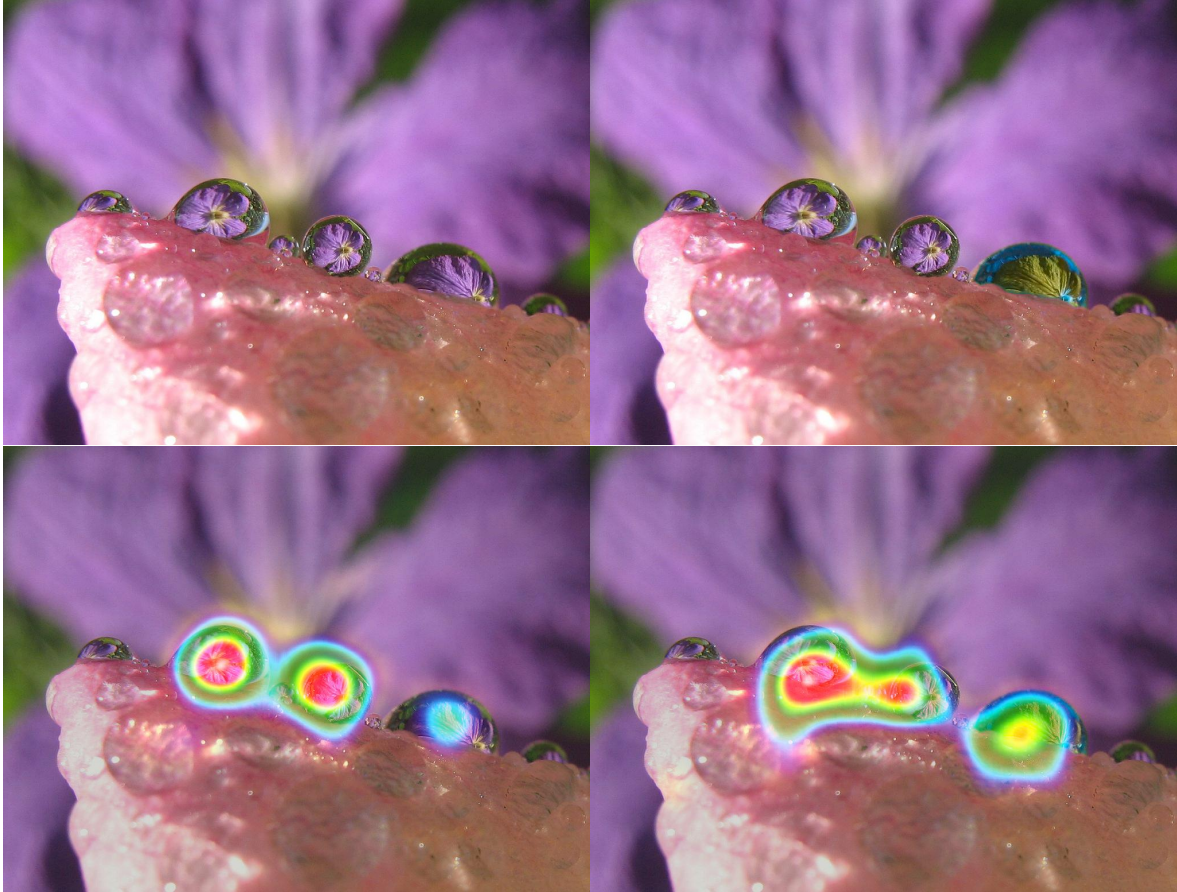
Figure 3.7: Examples of original and modified images used in our experiments with heatmap visualization of fixations. Each 2-by-2 array that corresponds to an image is organized as follows: original image (top-left); modified image (top-right); heatmap of fixations for original image (bottom-left); heatmap of fixations for modified image (bottom-right).

of 296 and 286 per viewer across all images) for the test and control groups, respectively. In addition to an overall increase in the total number of fixations on the ROI for the test group (by a factor of 2.79), a persistent increase in the lengths of these fixations is apparent. These results indicate that viewers are likely to fixate on the ROI of the modified images more often and for longer periods of time, compared to the ROI in the original image. Some examples of modified images are shown in Fig. 3.7 alongside fixation heatmaps that illustrate the shift in attention towards the modified ROIs.

3.4.2 Naturalness Evaluation

To evaluate the naturalness of our modified images, we conducted a subjective comparison between our images and those modified with other attention retargeting methods. In
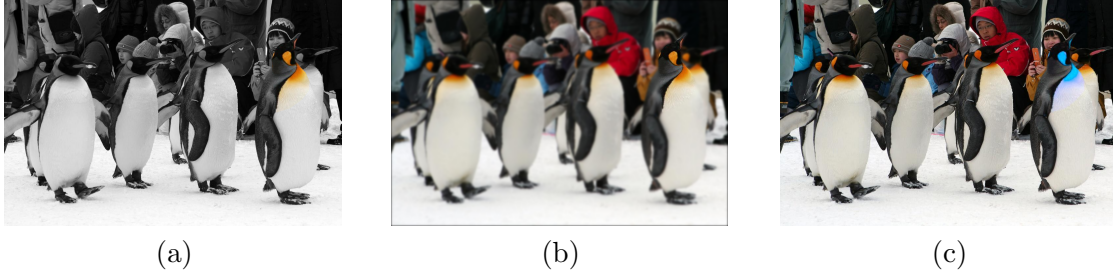
Figure 3.8: Examples of attention retargeting to the rightmost penguin's head (a) Monochrome effect (b) Gaussian blurring (c) Proposed method.

each trial, two images were displayed side-by-side for 4 seconds on a mid-gray background. One of the images was obtained by our method and the other was obtained using a different method. Each image was equally likely to appear on either side, i.e., 17 of our images randomly appeared on the left, while the other 17 randomly appeared on the right. The images were shown on a $27''$ ASUS LCD monitor with a resolution of 1920×1080. Ideally, we would have liked to compare our method against another state-of-the-art color-based saliency manipulation method. We contacted the authors of [40], [12], and [29], but were not able to obtain implementations of their methods. Hence, we chose a comparison similar to the one made in [29], against 1) "Monochrome" effect, in which the entire image is converted to grayscale, except for the ROI and 2) "Gaussian blurring," in which a Gaussian filter ($\sigma = 3$) is applied to the entire image, except for the ROI. Examples are shown in Fig. 3.8.

For every trial, participants were asked "which of the two images looks more natural?" in a Two Alternative Forced Choice (2AFC) task [35]. In a 2AFC, the participant is forced to make a decision between the two available images regardless of how uncertain they are. This methodology is more robust to measurement noise than scale-based quality ratings, such as the 5-point scale used in the naturalness evaluation found in [29]. This is because the naturalness of an image, which can be defined as "the degree of correspondence between the visual representation of the image and knowledge of reality as stored in memory," [19] is a fairly subjective and abstract concept. Mapping different degrees of naturalness to a fixed set of numbers cannot be accurately done unless the participant is trained beforehand.

Our naturalness evaluation was conducted separately from the eye-tracker tests, with a total of 24 participants. A two-sided chi-square χ^2 test was performed on the voting results of each image (Table 3.1) under the null hypothesis that both images are equally natural.

Under the null hypothesis, we expect an average of 12 votes for our image and 12 votes for the opposing image. The null hypothesis is rejected if $p < 0.05$, which indicates that there is a significant difference in terms of naturalness between the two images. In our comparison against the monochrome effect, the null hypothesis is rejected for 19 images, all of which have a significantly higher number votes in favor of our method. Against Gaussian blurring, the null hypothesis is rejected for 16 images, 15 of which favor our method, with only 1 in favor of the Gaussian blur. These results suggests that our method produces images that are equally or more natural looking compared to the monochrome effect, and equally or more natural most of the time compared to Gaussian blurring.

3.5   Conclusion

Systematic manipulation of color to guide visual attention may prove challenging, especially with the burden of maintaining the natural look of the original image. In this work, we developed a saliency manipulation method that modifies hue while keeping intensity and chromaticity constant. We describe the hue content of a ROI and its surroundings using a polar representation of a perceptually uniform color space, which allows us to easily determine the optimal hue adjustment to maximize the dissimilarity between the ROI and its surroundings. In addition to being simple and effective, the methodology makes it clear why saliency is changed as intended. We apply our method to maximize the saliency of selected ROIs in a set of natural images and confirm its effectiveness in guiding attention through eye-tracking. The naturalness of the results was evaluated in a separate set of subjective experiments.

Table 3.1: A comparison of the proposed method to monochrome effect and Gaussian blurring based on the voting results of our subjective experiments.

Image	Monochrome	Proposed	p-value	Blur	Proposed	p-value
Uniforms	5	19	0.0043	4	20	0.011
adsaliency09	8	16	0.1025	3	21	0.0002
adsaliency14	14	10	0.4142	11	13	0.6831
adsaliency21	4	20	0.0011	2	22	0.0000
adsaliency26	9	15	0.2207	3	21	0.0002
adsaliency27	10	14	0.4142	11	13	0.6831
adsaliency31	3	21	0.0002	5	19	0.0043
i1011319098	6	18	0.0143	12	12	1.0000
i1031604161	4	20	0.0011	9	15	0.2207
i1066946823	11	13	0.6831	17	7	0.0412
i1185710392	7	17	0.0412	2	22	0.0000
i1235260142	11	13	0.6831	9	15	0.2207
i1267668332	4	20	0.0011	10	14	0.4142
i1295408832	5	19	0.0043	6	18	0.0143
i1429029695	7	17	0.0412	6	18	0.0143
i1540552783	9	15	0.2207	10	14	0.4142
i169636965	11	13	0.6831	12	12	1.0000
i1795912442	8	16	0.1025	13	11	0.6831
i1870142757	7	17	0.0412	7	17	0.0412
i1893435749	1	23	0.0000	8	16	0.1025
i2132553812	7	17	0.0412	5	19	0.0043
i2145105890	5	19	0.0043	1	23	0.0000
i2145575787	3	21	0.0002	2	22	0.0000
i2186383189	12	12	1.0000	14	10	0.4142
i2200082985	8	16	0.1025	15	9	0.2207
i2240569900	7	17	0.0412	10	14	0.4142
i2259160448	6	18	0.0143	4	20	0.0011
i2263931014	8	16	0.1025	12	12	1.0000
i2273330095	9	15	0.2207	9	15	0.2207
i2288435981	3	21	0.0002	5	19	0.0043
i436895919	11	13	0.6831	8	16	0.1025
i441014928	1	23	0.0000	0	24	0.0000
ioffice_indoor	7	17	0.0412	13	11	0.6831
toronto84	12	12	1.0000	12	12	1.0000

Chapter 4

Subliminal Orienting

Subtlety is an important aspect of attention retargeting. Modifications that degrade an image so severely that they detract from the viewing experience are of limited practical use, regardless of whether they guide a viewer's attention as intended. It would be of great value to be able to guide attention in the least intrusive manner possible. Ideally, the modifications wouldn't be perceivable at all, thus guiding attention subliminally. Retargeting attention in images and videos without any perceivable alteration from the original stimuli may seem like an implausible concept. However, it is a topic of research in neuroscience and psychology that has received a fair bit of interest [28].

Although the studies reviewed in [28] demonstrate that it is possible to guide attention without awareness, they all use synthetic stimuli, and they use reaction time as a proxy for attention. It is uncertain whether such effects on attention would be observed if similar cues were to be applied in natural images and videos. Since reaction-time target detection tasks are less suitable for natural stimuli, investigation of subliminal orienting in natural images and videos would require eye-tracking as a more direct measure of visual attention. The only study to have done this is the work by Huang *et al.* [15], who briefly displayed a circular blob prior to displaying an image in an attempt to subconsciously draw attention toward the blob's location. The work in this chapter builds upon that of Cheadle *et al.* [3], which showed that subliminal flicker is capable of drawing attention in reaction-time tasks with simple synthetic stimuli. We test whether the same effect is observed in natural stimuli. The material in this chapter has been presented in [25].
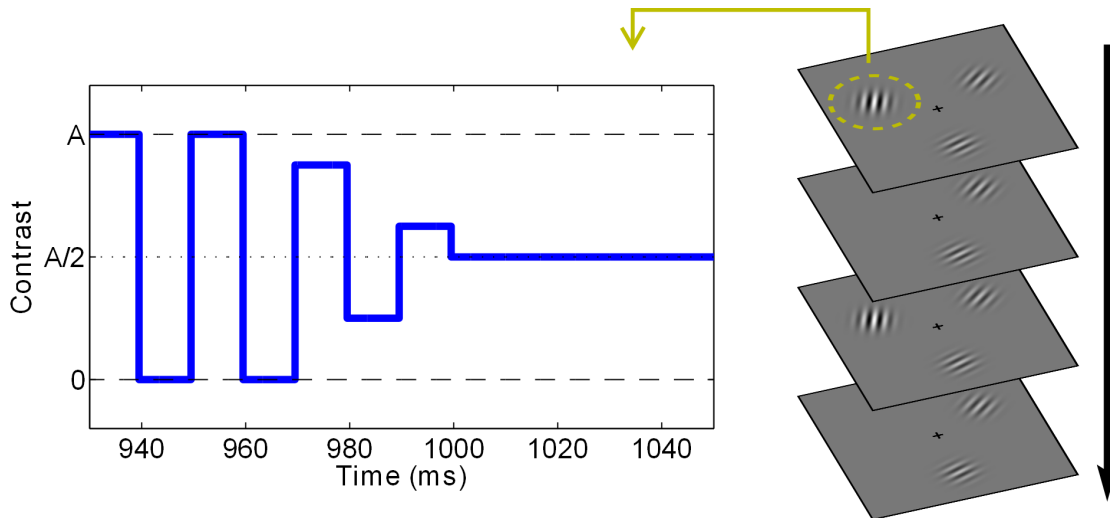
Figure 4.1: Experimental setup used in [3] to investigate the effects of subliminal flicker on attention.

4.1 Subliminal Flicker

A study by Cheadle *et al.* measured the response times of participants in a 3AFC task to investigate subliminal flicker as a cue for orienting attention [3]. The display used in their experiment consists of three Gabor patterns, equally spaced on an invisible circle around the black fixation cross at the center, as shown in Fig. 4.1. In a subset of trials, one of the three Gabor patterns flickered—its contrast alternating from maximum to minimum—at a frequency of 50 Hz (corresponding to a 100 Hz refresh rate). The flicker is no longer discernable at such a high frequency and the flickering Gabor patch appears identical to the other two. Afterwards, the spatial frequency of one of three Gabor patches was changed. Participants were told to locate this sudden change in spatial frequency as fast as possible. Their reaction times were found to be, on average, 15 ms faster in cases where the flicker is presented in the same location as the subsequent change. This implies that the flicker, though imperceptible, drew the subject's attention to its location, notifying them of the change ahead of time.

These findings are important because they suggest that attention can be retargeted without perceivable modifications. However, it has not been shown that the attentional orienting effects of subliminal flicker occur in natural stimuli. To investigate this, we apply the same type of flicker used in [3] to selected regions in a set of natural images and measure

38

(a)   (b)

Figure 4.2: Example image with (a) low-contrast ROI and (b) high-contrast ROI. These images are shown sucessively at 100 Hz to create a 50 Hz flicker. The ROI here is the container with the orange top in the lower left part of the image.

the resulting changes in attention through eye-tracking. The following section outlines our experiment.

4.2   Experiment on Natural Images

We selected 25 natural images from a dataset by Judd *et al.* [20] that varied in complexity from simple landscape images to scenes containing complex salient features, such as faces and text. Within each image, we selected ROIs away from those that drew a large number of fixations in eye-tracking data included in this dataset. ROIs were chosen to either be on the left or right side of the image with no ambiguity. The ROI in each image was modified to create two different versions of each image: one with a high contrast ROI and one with a low contrast, as shown in Fig. 4.2. This was done by maximizing the variation in contrast without forcing values outside the dynamic range, meanwhile ensuring that the time-averaged contrast was equivalent to that of the original ROI. When these two modified images are alternately shown in rapid succession, henceforth denoted as "flicker," the ROI appears identical to the original image. The code used to generate these stimuli is available at www.sfu.ca/~ibajic/software.html.
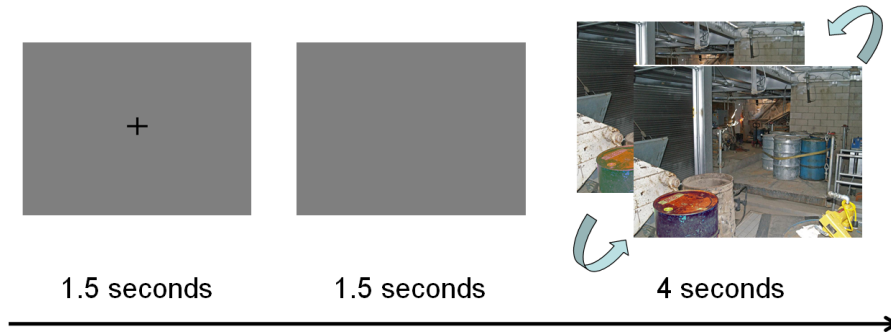
39

Figure 4.3: Format of the eye-tracking test for flickering images.

4.2.1 Eye-Tracking Test

Participants wore a head-mounted Locarna Pt-Mini eye-tracker, while maintaining a viewing distance of 70 cm via a chin rest. The eye-tracker was calibrated prior to each test using 9 reference points displayed on the monitor used for testing. The localization task was performed after the eye-tracking test, at a viewing distance of 70 cm without the chin rest. All tests were shown on a ViewSonic Graphics Series G220f CRT monitor with a resolution of 1024×768 at a refresh rate of 100 Hz in a room with an ambient light of 200 lx.

Each stimulus was shown for 4 seconds on a blank (mid-level grey) screen, followed by a black crosshair displayed for 1.5 seconds at the center of a blank screen. Participants were told to fixate on the crosshair whenever it appeared, which helped us judge the accuracy of the eye-tracker throughout the test. A blank screen was displayed for 1.5 seconds after the crosshair was shown to prevent any center bias for the upcoming stimulus. The test group of subjects was shown the flicker images, one with high-contrast ROI and one with low-contrast ROI, alternating at a frequency of 50 Hz, throughout their entire 4-second duration. The control group was shown the original static images. All stimuli appeared in random order in each test.

Eye-tracking data was collected from 22 nonexpert participants (17 male, 5 female), between the ages of 18-30, with either normal vision or corrected-to-normal vision via contact lenses. A control group of 11 participants viewed the original, static images. The other 11 participants (test group) viewed the flickering images.

We define a fixation as a set of consecutive gaze data points that lie within $1°$ of visual angle from each other for a minimum of 0.167 seconds [6]. A fixation is considered to be on the ROI if it is located within the mask of the ROI after it has been dilated with a
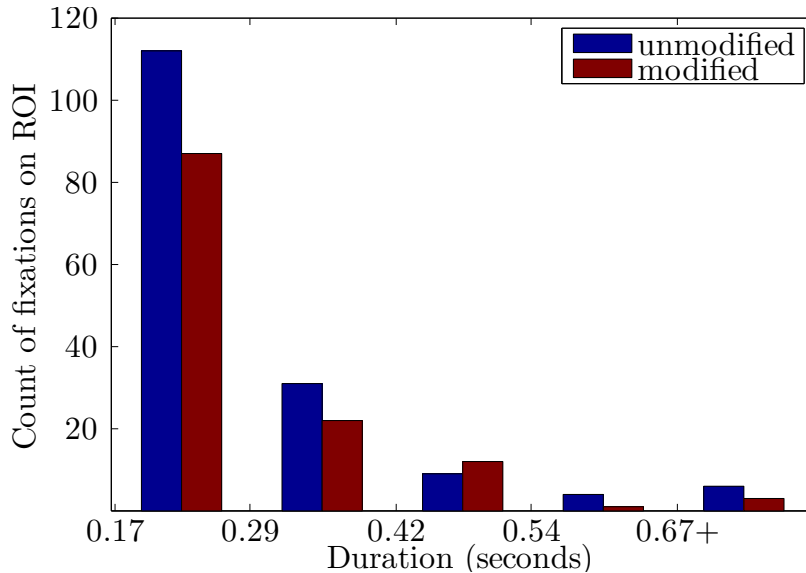
Figure 4.4: Histograms of the duration of fixations within the ROI.

circular disk with a radius of 1° of visual angle to account for possible eye-tracking errors. A histogram count of the duration of fixations on the ROI for the control group (blue) and the test group (red) is shown in Fig. 4.4. The number of fixations on the ROI were 125 and 162 out of a total of 2226 and 2403 across all images for the test and control groups, respectively. Thus, we observe a decrease in the overall number of fixations on the ROI in the test group relative to the control by a factor of 0.77. Heatmap visualizations of the fixation data for all images are available with the supplementary material at www.sfu.ca/~ibajic/software.html.

4.2.2 Localization Task

Participants were provided with written instructions, which asked them to locate a flicker within each image. Each stimulus was shown for 4 seconds in random order, followed by a 1.5-second blank screen. During this time, participants were asked to decide whether the flicker was on the left or right side of the image via arrow key inputs on a keyboard. An input could be changed at any time before the next stimulus appeared. If participants chose incorrectly, a beeping sound was played to help motivate them to improve their performance.

The localization task was performed by all 22 participants after they completed the eye-tracking test. If the flicker were truly imperceptible, we would expect the average localization accuracy over many subjects and images to be 0.5, called the chance level. However, our set of images is relatively small and side-biased because out of 25 images,

15 have ROI on the left and only 10 on the right. To remove this bias, we consider all possible datasets that can be formed by removing 5 images with ROI on the left. In total, there are $\binom{15}{5} = 3003$ possible unbiased datasets that can be formed by removing 5 left-ROI images. We compute the 95% confidence interval for the mean accuracy in each one of these unbiased datasets and obtain the average of all left endpoints and all right endpoints. The final side-bias-corrected confidence interval is $[0.50, 0.63]$. Since the chance level lies on the boundary of this interval, the data suggests that the flicker is subliminal at the 95% confidence level.

4.3 Conclusion

The results of this preliminary study seem to indicate that subliminal flicker, on average, works to repel attention away, rather than draw it to its location in natural images. The histogram in Fig. 4.4 shows a decrease in the overall number of fixations on the ROI of the flickering images, relative to the original flicker-free images. This effect can be observed in various fixation heatmaps for each image as well. Fixations appear to be more scattered in the original static images. Among the test images, there were a few cases that showed increased concentration of fixations near the flickered ROI, however, their number was small compared to those images that showed the opposite effect. The impact of image content on the ability of subliminal flicker to draw attention is an issue that requires further investigation.

A more detailed analysis is necessary to determine whether the fixations between the test group and the control group are significantly different, and until signifance is confirmed, further speculation as to the underlying causes of these differences cannot be made. Based on our current evidence, it appears that on average, subliminal flicker does not attract attention in natural images.

Chapter 5

Conclusions

5.1 Summary of Contributions

This thesis presents an introduction to attention retargeting and its connection to visual saliency. We define attention retargeting as a saliency inversion problem and specify the main challenges involved. Some general approaches to this problem are presented with examples from existing work in an attempt to arrange these works within a unified framework, which was previously nonexistent. Though we do not claim that our problem definition and subsequent analysis is definitive, we hope that the foundation laid here will help provide a new perspective on the existing methodologies.

We propose two novel attention retargeting methods to predict the extent to which a viewer's gaze will be drawn towards a ROI after that region is modified. The first operates on orientation—a previously unexplored visual feature for attention retargeting—by rotating the ROI. The second operates on the color of the ROI. Both methods map relative ROI saliency to their respective visual modifications in an efficient manner. We choose the modification that maximizes the relative saliency of the ROI for our evaluations. In addition, the methodology makes it clear why this modification maximizes relative saliency.

There are two issues to consider when evaluating attention retargeting models: 1) Does the retargeting guide attention as intended? 2) Does the retargeting harm image naturalness? We used eye-tracking in our evaluations as a direct measure of visual attention. A group of participants views the modified dataset, while a control group views the original dataset. To demonstrate an increase of attention on the ROI when viewing the modified dataset, we take into account both the number of fixations on the ROI, as well as their

lengths. An increase in these two attributes over the control group reliably indicates that viewers tend to be drawn to the ROI of the modified images more often. Heatmap visualizations of fixations are also provided to illustrate the general gaze patterns among all viewers of a particular group. As a reliable measure of naturalness, we conduct subjective tests where participants are forced to decide which image looks better in a comparison of two images, both of which are modified by two different retargeting algorithms. We hope that our evaluation methodology serves as a good model for future research in the field.

We also analyzed the plausibility of subliminal flicker as a means for drawing attention in natural images without the viewer's knowledge. We alternated the contrast from high to low at a frequency of 50 Hz within selected regions in a set of natural images. Eye-tracking data was collected on a group of participants that viewed the flickering images and another group that viewed the original static images. A subsequent localization task was used to determine whether the flicker was truly subliminal. A comparison of the eye-tracking data between the two groups indicated that subliminal flicker may, on average, repel attention rather than attract it.

5.2   Future Work

As a relatively unexplored topic, attention retargeting presents a wide array of possibilities for improvement. Though we may still be far from a saliency inversion algorithm capable of projecting changes from a target saliency map onto a desired set of features in an image with minimal harm to its naturalness, several intermediate steps can be taken. The straightforward approach of reverse-engineering existing saliency models may provide valuable insight into attention retargeting. A reliable measure of naturalness, especially for ROI-based modifications, may be of critical importance. Methods to predict how changes in a combination of image features affect saliency can be a crucial development. Furthermore, it may also be worthwhile to investigate the differences in gaze patterns on retargeted images between specific demographics, e.g., a comparison between men and women, teenagers and seniors, artists and accountants, etc. Finally, the plausibility of subliminal orienting in natural images and video remains an open issue. We hope that our work sparks interest in attention retargeting and motivates others to contribute.

Bibliography

[1] A. Borji and L. Itti. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):185–207, Jan. 2013. 4

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, Nov. 1986. 18

[3] S. W. Cheadle, A. Parton, H. J. Müller, and M. Usher. Subliminal gamma flicker draws attention even in the absence of transition-flash cues. *J. Neurophysiol.*, 105(2):827–833, Feb. 2011. xi, 37, 38

[4] A. Criminisi, P. Prezes, and K. Toyama. Object removal by exemplar-based inpainting. In *Proc. IEEE CVPR'03*, Madison, WI, Jun. 2009. 18

[5] R. Dayhot. Statistical hough transform. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(8):1502–1509, Aug. 2009. 15, 16

[6] A. Duchowski. *Eye Tracking Methodology: Theory and Practice.* Springer, 2nd edition, 2007. 19, 30, 40

[7] R. O. Dude and P. E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Comm. ACM*, 15:11–15, Jan. 1972. 14

[8] S. Fuller, Y. Park, and M. Carrasco. Cue contrast modulates the effects of exogenous attention on appearance. *Vision Research*, 49(14):1825–1837, 2009.

[9] A. Garcia-Diaz, X. R. Fdez-Vidal, X. M. Pardo, and R. Dosil. Saliency from hierarchical adaptation through decorrelation and variance normalization. *Image Vis. Comput.*, 30(1):51–64, Jan. 2012. 3

[10] T. Gevers and H. Stokman. Robust histogram construction from color invariants for object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):113–118, Jan. 2004. 26

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *Proc. IEEE CVPR'10*, pages 2376–2383, Jun. 2010. 3

[12] A. Hagiwara, A. Sugimoto, and K. Kawamoto. Saliency-based image editing for guiding visual attention. In *Proc. PETMEI'11*, pages 43–48, Sep. 2011. 9, 24, 34

[13] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *Advances In Neural Information Processing Systems 19*, pages 545–552. MIT Press, 2007. 3

[14] P. Hough. Methods and means for recognizing complex patterns. US Patent 3,069.654, 1962. 14

[15] T. H. Huang, Y. H. Yang, H. I. Liao, S. L. Yeh, and H. H. Chen. Directing visual attention by subliminal cues. In *Proc. IEEE ICIP'12*, pages 1081–1084, Oct. 2012. 37

[16] L. Itti and C. Koch. Computational modeling of visual attention. *Nature Rev. Neurosci.*, 2(3):194–203, Mar. 2001. 2

[17] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, Nov. 1998. x, 3, 4, 9

[18] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989. 25

[19] T. J. W. M. Janssen and F. J. J. Blommaert. Image quality semantics. *Journal of Imaging Science and Technology*, 41(5):555–560, 1997. 34

[20] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *ICCV*, pages 2106–2113, 2009. 18, 29, 39

[21] W. Kim, C. Jung, and C. Kim. Spatiotemporal saliency detection and its applications in static and dynamic scenes. *IEEE Trans. Circuits and Systems for Video Technology*, 21(4):446–456, Apr. 2011. 3

[22] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statistics*, 22(1):79–86, 1951. 17

[23] V. A. Mateescu and I. V. Bajić. Guiding visual attention by manipulating orientation in images. In *Proc. IEEE ICME'13*, Jul. 2013. 13

[24] V. A. Mateescu and I. V. Bajić. Attention retargeting by color manipulation in images. In *ACM Multimedia PIVP*, pages 15–20, Nov. 2014. 23

[25] V. A. Mateescu and I. V. Bajić. Can subliminal flicker guide attention in natural images? In *ACM Multimedia PIVP*, pages 33–34, Nov. 2014. 37

[26] V. A. Mateescu and I. V. Bajić. Visual attention retargeting. Submitted for publication in *IEEE Multimedia Mag.*, Oct. 2014.

[27] M. Mulckhuyse, D. Talsma, and J. Theeuwes. Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Visual Cognition*, 15(7):779–788, 2007.

[28] M. Mulckhuyse and J. Theeuwes. Unconscious attentional orienting to exogenous cues: A review of the literature. *Acta Psychologica*, 134(3):299–309, 2010. 37

[29] T. V. Nguyen, B. Ni, H. Liu, W. Xia, J. Luo, M. Kankanhalli, and S. Yan. Image re-attentionizing. *IEEE Trans. Multimedia*, 15(8):1910–1919, Dec. 2013. 11, 29, 30, 34

[30] Niebur E. Parkhurst, D. J. Texture contrast attracts overt visual attention in natural scenes. *Eur. J. Neurosci.*, 19(3):783–789, 2013.

[31] H. J. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *The Journal of Vision*, 9(12):1–12, Nov. 2009. 3

[32] G. Sharma, editor. *Digital Color Imaging Handbook*. CRC Press, 2003. 25

[33] S. L. Su, F. Durand, and M. Agrawala. De-emphasis of distracting image regions using texture power maps. In *Proc. Fourth IEEE Int'l Workshop Texture Analysis and Synthesis (Texture '05)*, pages 119–124, 2005. 9

[34] J. R. Taylor. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books, 2nd edition, 1982. 26

[35] M. Taylor and C. Creelman. Pest: Efficient estimates on probability functions. *Journal of Acoustical Society of America*, 41:782–787, 1967. 34

[36] A. M. Treisman and G. Gelade. A feature-integration theory of attention. *Cogn. Psychol*, 12:97–136, 1980. 2

[37] J. K. Tsotsos. Analyzing vision at the complexity level. *Brain Behav. Sci.*, 13:423–469, 2001. 2

[38] J. M. Wolfe and T. S. Horowitz. What attributes guide the deployment of visual attention and how do they do it? *Nature Rev. Neurosci.*, 5:1–7, 2004. 2

[39] L. Wong and K. Low. Saliency-enhanced image aesthetics class prediction. In *Proc. IEEE ICIP09*, pages 997–1000, Nov. 2009.

[40] L. Wong and K. Low. Saliency retargeting: An approach to enhance image aesthetics. In *IEEE Workshop on Applicat. of Comput. Vision (WACV'11)*, pages 73–80, Kona, HI, Jan. 2011. 6, 24, 34

[41] YouTube. Guiding visual attention by manipulating orientation in images gaze data visualizations. *YouTube.com*. [Online] Available: http://www.youtube.com/watch?v=9cDOSO7MXXc. [Accessed: April 28, 2012]. 20