Using the Earth Mover's Distance for Perceptually Meaningful Visual Saliency

by

Logan Dunbar

Department of Electrical Engineering

University of Cape Town

A dissertation submitted in fulfilment of the requirements for the degree of

Master of Science in Engineering

June 2015

Supervisor:

Prof. Fred Nicolls

Declaration

I, Logan Dunbar, confirm that:

- This work was completed while registered for a Master of Science in Engineering degree at the University of Cape Town.
- This work has not previously, in its entirety or in part, been submitted for obtaining any other qualification at this, or any other, university.
- The external sources of information used herein have been clearly attributed.
- I know the meaning of plagiarism and declare that all the work in this document, save for that which is properly acknowledged, is my own.

Signed:

Date:

Abstract

Visual saliency is one of the mechanisms that guide our visual attention, or where we look. This topic has seen a lot of research in recent years, starting with biologically-inspired models, followed by the information-theoretic and recently statistical-based models. This dissertation looks at a state-of-the-art statistical model and studies what effects the histogram construction method and histogram distance measures have on detecting saliency.

Equi-width histograms, which have constant bin size, equi-depth histograms, which have constant density per bin, and diagonal histograms, whose bin widths are determined from constant diagonal portions of the empirical cumulative distribution function (ecdf), are used to calculate saliency scores on a publicly available dataset. Cross-bin distances are introduced and compared with the currently employed bin-to-bin distances by calculating saliency scores on the same dataset. An exhaustive experiment with combinations of all histogram construction methods and histogram distance measures is performed.

It was discovered that using the equi-depth histogram is able to improve various saliency metrics. It is also shown that employing cross-bin histogram distances improves the contrast of the resulting saliency maps, making them more perceptually meaningful but lowering their saliency scores in the process.

A novel improvement is made to the model which removes the implicit center bias, which also generates more perceptually meaningful saliency maps but lowers saliency scores. A new scoring method is proposed which aims to deal with the perceptual and scoring disparities.

Acknowledgements

Thank you to everyone who had to put up with my antics while I finished this dissertation. It was a long and trying road, but I am incredibly lucky and happy to have had the opportunity to travel it with you all.

To Mom, without you this would not have seen the light of day. You have been an ever present source of support and love for me, and I thank you with all my heart.

To Dad, sorry you aren't here to see this, but I know you would be proud.

To Lauren and James, thanks for always sticking by me, lifting my spirits, and egging me on to finish. You are the best siblings a man can hope for.

Contents

List of Figures

List of Tables

Chapter 1
Introduction

The natural world is an extremely complex and intricate system. A 250ml glass of water contains roughly 8.25×10^{24} H_2O molecules[1] all exerting forces on each other, the container and the environment. That is about $1000\times$ more molecules than there are cups of water in all the water on earth! To fully characterise this system we would need to be able to specify every single force and action of every single molecule, and even then that does not include the details of the quarks and leptons at an even smaller scale. Yet, through all this complexity, we humans and animals alike have developed an ability to distil the essence of the world around us which enables us to understand it and interact with it. If we were to knock over the glass we could, with moderate accuracy, predict where the water might flow. The explosion in complexity as we look at the system through the molecular lens is called the curse of dimensionality [6], which luckily nature has provided us and the animal kingdom a powerful tool to overcome: abstraction. Abstraction allows us to look at the world through generalisations, to change the scale of the world we are looking at. Rather than attempting to identify and model each molecule individually, we are able to identify "water" as a single concept and assign properties to it, such as its viscosity or its ability to take on a container's shape. It might not be the whole picture, but it is certainly enough to allow us to use it for our survival and competitive advantage.

Survival of the fittest requires that a creature have some trait or ability which allows it to make better use of its environment which would allow it to propagate its species forward through time. Over the course of the world's evolution nature has designed many solutions for interacting with the environment around us, most notably sight. The ability to turn electromagnetic waves into electrical and chemical signals that represent the world, otherwise known as sight, was obviously an incredibly important advantage, judging by the ubiquity of the eye in nature. To be able to interact with the world the eye could try and identify every molecule in every object, but the amount of processing power required for such a task would be enormous. Our brains exist

[1] For water at 20°C and 1atm: 250ml × 0.9882g/ml [26] × 6.0221×10^{23}molecules/mol ÷ 18.01528g/mol [47] = 8.25×10^{24} H_2O molecules.

in a finite space and require energy to function, meaning the processing power at our disposal is a finite resource. To deal with these limitations, the eye developed its own abstractions. There are approximately 92 million rods and 5 million cones in the average human eye, but only about 1 million connections leading from the eye to the visual cortex [12]. Even then, the information is not sent directly as it is received; rather it is preprocessed into roughly 10-12 channels consisting of abstracted information such as edges, motion and large areas of uniform colour in the scene. The visual system does something we refer to as sparse coding. It finds the minimum amount of information that adequately allows the brain to reconstruct the scene which gives the creature the best possible chance of survival [29].

To keep the entire scene in front of us in focus would require a tremendous amount of data capture, as well as processing. We have developed a mechanism called visual attention, otherwise known as selective focus, to make this process more efficient. Rather than attempting to keep the entire scene in focus, a "spotlight" of higher resolution imagery is allowed to scan the scene in rapid succession, called saccades [37]. The assumption is that the environment does not change more rapidly than we can update our representation of it, which in most situations is reasonable. This ability allows us to selectively view the scene, allocating our finite processing power only to the parts of the scene that require it or that we deem important.

There are certain triggers that guide this attentional mechanism, the most popular theory being the corresponding bottom-up and top-down attentional mechanisms [23, 27, 40, 44]. The theory suggests that we have two streams of attention, one being a highly parallel feature and stimulus driven mechanism, the other being a much higher-level goal-driven mechanism. These two systems work in tandem to decide what part of the visual field should be allocated the spotlight of processing power. Sharp edges, abrupt colour changes, repeating patterns, moving objects and even seemingly higher-level concepts like faces can trigger the bottom-up mechanism to indicate interesting content [25]. Top-down goals such as looking for words, or a specific colour or face in a crowd, would basically prime the bottom-up process to only respond to those cues, overriding its default behaviour. These processes result in certain parts of the world standing out relative to their surroundings and this perceptual quality is known as visual saliency.

In recent years, researchers have developed models for saliency which have moved from the biologically-plausible models [22, 23, 27, 34, 40] into the information-theoretic [9, 15], statistical [33, 50] and transform-based [18, 31] models, each providing new and unique understandings of the way visual saliency works. Biologically-based saliency is a great place to start, due to it being successfully implemented in nature and therefore readily available to be studied, but perhaps through biological limitations or evolutionary pressures the current systems are sub optimal.

1.1 Objectives and Hypotheses

The statistical models have recently shown good performance at detecting visual saliency and predicting where humans might look, which is why this dissertation focuses on exploring these models and attempts to discover the effects certain assumptions and design decisions have on the resulting saliency predictions.

The model presented by Liu et al. [33] makes extensive use of histograms to calculate the saliency in an image, where a histogram is a mathematical tool that provides an estimate of a data distribution. The authors use equi-width histograms that partition the colour space, in which the image resides, into equal-sized bins. However, as shown by Piatetsky-Shapiro and Connell [36], equi-depth histograms that vary the bin width to maintain a constant number of data points per bin achieve lower estimation errors. Further, Denby and Mallows [13] show that equi-width histograms are unable to capture spikes in the distribution and that equi-depth histograms smooth the low-density regions of the distribution. They introduce the new diagonal histogram (d-hist) which is a trade-off between the two. The first objective of this dissertation is to discover what effect the histogram construction method has on detecting visual saliency. It is hypothesised that the variable width equi-depth and d-hist construction methods will be better able to estimate the data distribution in the images, and will therefore allow better detection of visual saliency.

Histogram distances also play a crucial role in the model of Liu et al. [33]. The authors use histogram distances to compute saliency values based on the assumption that salient regions generally show contrast with the surrounding background regions as well as the assumption that salient colour distributions are more sparsely distributed around the image than background colour distributions. They use a combination of bin-to-bin distances and cross-bin distances to compute the saliency values. Bin-to-bin distances are highly susceptible to binning effects, which occur when the bin edge incorrectly partitions related data, and histogram shifting, which easily occurs due to lighting and shading in the image. The second objective of this dissertation is to identify whether the histogram distance used has an impact on the performance of detecting visual saliency. It is hypothesised that using cross-bin distances which negate the susceptibility to binning effects and histogram shifts will improve the ability to generate meaningful saliency values.

The third objective is to identify whether the methods of histogram construction and histogram distances depend on one another, and if there is a combination that provides the best result. It is hypothesised that combining the variable width histogram construction methods with the cross-bin histogram distances will provide a jointly improved ability to detect visual saliency.

An issue present in most saliency scoring methods as well as saliency models is that of center bias. Photographers generally place objects of interest in the center of the frame and datasets generally have participants fixate on the center of the frame before being exposed to stimuli. This leads to inflated scores when the saliency model has a built-in center bias, as does the model of Liu et al. [33]. The fourth and final objective of this dissertation is to attempt to remove the center bias from the model. It is hypothesised that removing the center bias will lower the resulting saliency scores due to the center bias artificially inflating most scoring methods. However, it is also hypothesised that removing the center bias will improve the perceptual output of visual saliency because the center bias will no longer mask salient regions towards the edges of the image.

1.2 Road Map and Results

Chapter 2 provides the background information for the techniques and methods used in this dissertation. The history of histograms and their construction methods are discussed, along with the various histogram distances which will be used in the experiments. Related research and key concepts of saliency are covered, and the model from Liu et al. [33] that forms the basis of study for this dissertation is presented. The methods and formulae for scoring saliency, the dataset and the viewing tasks performed, which consist of three eye-tracking tasks with corresponding eye-tracking data and one explicit selection task with corresponding mouse-click location data, close out the chapter.

Due to the different strengths and weaknesses of the histogram construction methods, an experiment is presented in Chapter 3 that determines their effects on saliency detection. It is shown that equi-depth histograms improve performance on a number of saliency scoring metrics for the eye-tracking data, but lower performance for the explicit selection task. The equi-depth histogram's smoothing of the low-density, or salient, colours causes the saliency in the image to become more spread out. This favours the more random and haphazard eye-tracking data and not the more precise explicit selection data.

An experiment to test the effects of the choice of histogram distance is presented in Chapter 4. It is shown that cross-bin distances are detrimental to the saliency scores generated, but perceptually improve the saliency detection in the image. The disparity is due to bin-to-bin distances not sufficiently suppressing the background regions which, when combined with the eye-tracking data, generate higher scores but have perceptually less meaningful results. It is noted that the eye-tracking data is spread out and more randomly distributed around the image than the explicit selection data, and a new scoring method is designed based on this observation.

The combined effects of the histogram construction methods and histogram distances are tested for in Chapter 5. The results from Experiments 1 and 2 are reiterated and show that the equi-depth histograms and cross-bin distances can improve saliency detection perceptually, but do not align well with current scoring methods.

An experiment is performed and presented in Chapter 6 which removes the center bias from the model and compares saliency scores with the original method. It is shown that removing the center bias from the model drastically lowers most scores, apart from the scoring method which accounts for the center bias, but the resulting saliency values are more perceptually meaningful. This highlights the need for new scoring methodologies that better align with perception.

Chapter 7 concludes the dissertation with final thoughts and insights, and provides some avenues for future research.

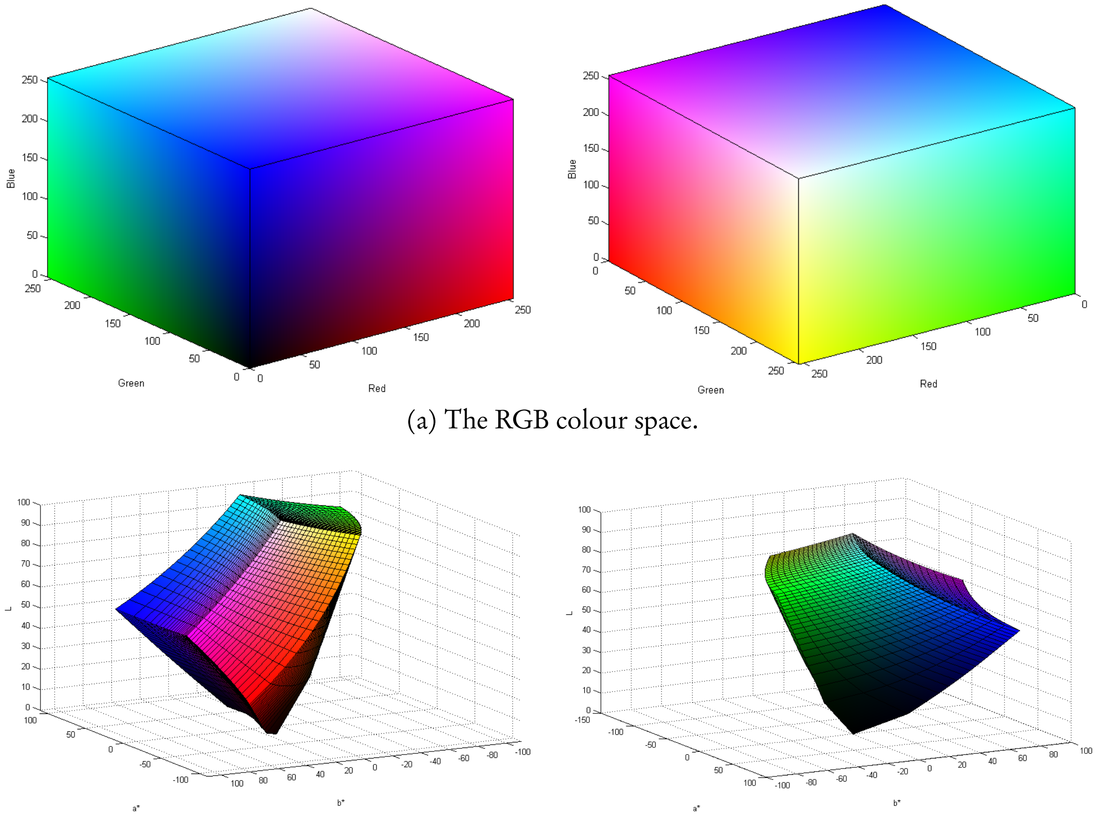
Chapter 2
Background Information

Saliency models and their subsequent scoring methods are often built up using existing mathematics and methods. This chapter serves to introduce and explain in detail how these concepts are applied when calculating and scoring saliency. Section §2.1 gives a brief history of how the concept of a colour space came about, and how it can be used to summarise an image's contents. The tool used to summarise the image is the histogram and its various construction methods are explored in detail in Section §2.2. It is possible and useful for computing saliency to compare these image summaries with one another by calculating distances between their histograms, and some histogram distance measures are shown and their strengths and weaknesses explained in Section §2.3. The concept of saliency, its origins, and its recent related literature is presented in Section §2.4. The saliency model defined in [33] is the foundation for the experiments in this dissertation, and is given a thorough working through in Section §2.4.1. Finally, Section §2.5 explains how to score saliency using four of the most common metrics, and Section §2.5.1 provides details of the dataset used in the experiments.

2.1 Colour Spaces

The perception of colour by humans is currently explained by two complementary theories. The Young-Helmholtz, or trichromatic, theory states that we have photoreceptive cells in our eyes that are sensitive to three wavelengths of the electromagnetic spectrum [42]. These wavelengths roughly correspond to the colours red ($\lambda = 560nm$), green ($\lambda = 530nm$) and blue ($\lambda = 420nm$). As one of the first validations of this theory, James Clark Maxwell took the first permanent colour photograph in 1861 by using three separate colour filters, namely red, green and violet-blue, which when viewed together in a dark room created the colour photograph[1]. This led to the notion of an additive colour space, known as the Red Green Blue (RGB) colour space, in which colours are reproduced by additively combining the three primary colours. This colour

[1] Photo viewable at `http://www.nationalgeographic.com/wallpaper/photography/photos/milestones-photography/color-tartan-ribbon/`.



(a) The RGB colour space.

(b) The RGB space embedded in the CIE $L^*a^*b^*$ space.

Figure 2.1.1: Representations of colour spaces in 3D.

space can be visualised or represented by a 3-dimensional space as seen in Figure 2.1.1a, where a vector of three values between 0 and 255 defines their additive contribution to the final colour.

The trichromatic theory could explain the physical mechanisms of colour, but it was noted that we never see reddish-green or blueish-yellow, and the theory could not explain the supposedly negative colour afterimages we perceive after overstimulation of the retina. Ewald Hering created the opponent process theory which postulates that colour perception is created by three opponent processes based on the primary red, green and blue receptors [5]. These opponent processes consist of opposing pairs of colours, namely white-black, red-green and blue-yellow. The responses from each channel in a pair are antagonistic to each other, which explains the inability to perceive reddish-green, it also explains the negative afterimages, as an overstimulation of red would create an afterimage of green; likewise overstimulation of blue would create an afterimage of yellow. It is thought that the overstimulation of one channel fatigues the neuronal firing of that channel, and when the stimulus is removed the opposing channel's signals are by comparison magnified.

The International Commission on Illumination (CIE)[2] was established in 1913, and is recognised by the International Organization for Standardization (ISO) as an international standardisation body dealing with all matters relating to the science and art of light and lighting, colour and vision, photobiology and image technology. In 1931 the CIE formally defined the CIE 1931 XYZ colour space, commonly referred to as CIE XYZ, based on experiments by William David Wright [48] and John Guild [16], whose purpose was to encompass all the colour sensations an average person can experience, and as such serves as a standard reference against which many other colour spaces are defined. CIE XYZ provides a mapping of physical light spectra onto human perceived colours, as represented by tristimulus values, which takes into account that humans overstate the contribution of green to the luminance or brightness of a colour, as well as some other adjustments due to the differences in our subjective perception of colour.

Although the CIE XYZ "master" space can represent every perceivable colour, it is not particularly perceptually uniform. Perceptual uniformity means a change of the same amount in the colour space should produce an equivalent perceptual change in the colour perceived. To overcome this shortfall, the CIE $L^*a^*b^*$ (CIELAB) colour space was defined in 1976 [21], which is a perceptually uniform opponent colour space with $\boldsymbol{L}$ for lightness and $\boldsymbol{a}$ and $\boldsymbol{b}$ for the colour opponent dimensions of red-green and yellow-blue respectively. This perceptual uniformity was achieved via a nonlinear transformation of the CIE XYZ colour space based on psychophysical experiments, as can be seen by embedding of the RGB space in the CIELAB space in Figure 2.1.1b. The beauty of having a perceptually uniform colour space is that it allows mathematical treatment of the differences between colours that correlate to our perception via the simple Euclidean distance between colours. It is thanks to this attractive property that the CIELAB colour space has become the de facto standard for colour computations in the computer vision community and is why it is chosen for the remainder of this dissertation.

There are many colour spaces defined, each having properties that make them attractive to the community that spurred their development. For example, the Hue-Saturation-Value (HSV) and Hue-Saturation-Lightness (HSL) colour spaces came about due to the difficulty of manipulating colours in RGB. To change the shade of red, one needs to update all three RGB values simultaneously, which can be tedious, unintuitive and error prone. The HSV and HSL spaces rearrange the RGB space into cylindrical

[2] `http://www.cie.co.at`.

spaces, hue being the angle around the colour wheel, the radius being the saturation of the specific hue, and the vertical axis being the lightness or brightness of the colour. Now changing the shade of red equates to adjusting its saturation or brightness level, creating a much more intuitive interface for graphical designers and digital artists. These spaces still suffer from the perceptual nonuniformity which makes them imprecise and inaccurate to work with computationally.

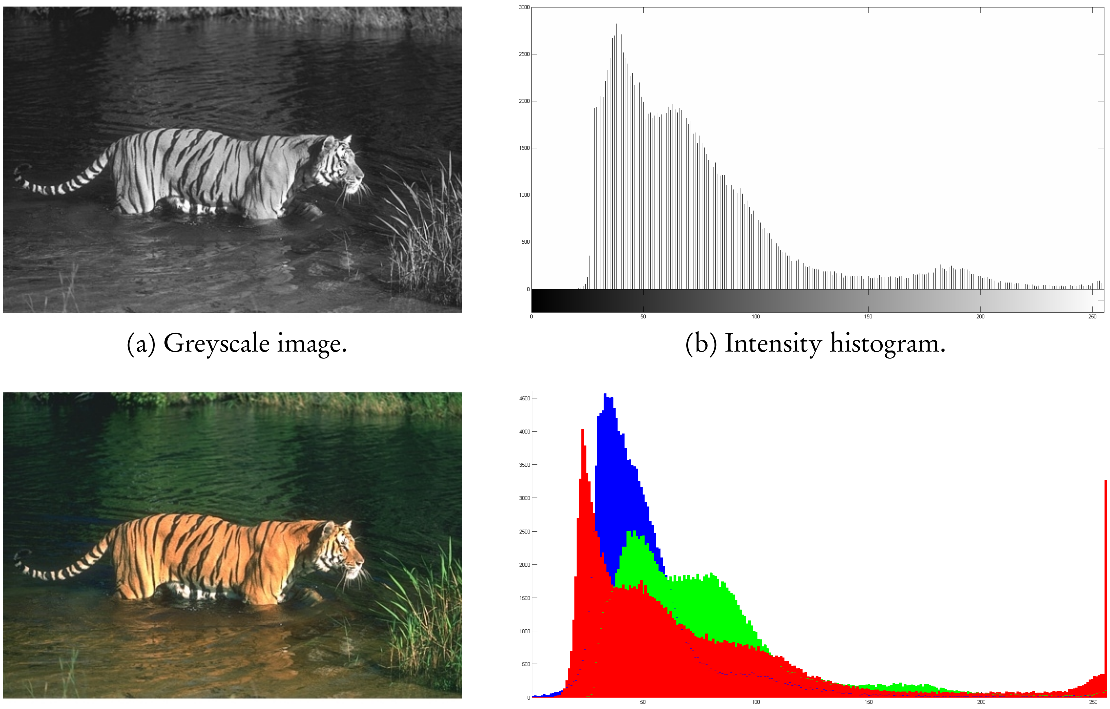
An image is made up of pixels, usually arranged in a rectangular grid, each of which has an associated colour triplet most commonly stored in the RGB format. An image can therefore be seen as a point cloud in the space of our choosing, which allows us to apply clustering and data analysis techniques to uncover hidden structures and manifolds in the image. Histograms are one such tool which allow us to achieve this.

2.2 Histograms

Histograms have their origins in the 19th century, introduced by Karl Pearson as a means for graphically representing frequency data. As a mathematical tool, a histogram provides an estimate of a data distribution, usually a discrete approximation of a continuous variable, but can be applied to discrete data equally well. It acts very much like a summary, exposing structures in the data as well as making computations tractable by approximating and condensing the data [19].

Histograms made their first commercial appearance in the 1980s in database query optimisers. By estimating how many records would need to be read or written (I/O), the query optimiser is able to reorder the query appropriately so as to minimise data I/O thereby drastically reducing query execution time. They then made their way into image processing and computer vision, being used to summarise the intensity information in an image, allowing characterisation of the content of the image, and allowing image enhancement by histogram equalisation. A colour histogram characterises the image in all three of its colour channels and adds further discriminatory and summarisation power. Examples of intensity and colour histograms are shown in Figure 2.2.1.

The most common and easy to implement histogram is the equi-width histogram. For the 1-dimensional case it is constructed by partitioning a region of the data space into equal width partitions, called bins, and then counting the number of data points that lie within each bin's edges. There are two main variables when dealing with histograms, total number of bins n, and bin width k. In the equi-width histogram n and k are related by $k = \frac{upper - lower}{n}$ where $upper$ and $lower$ are the upper and lower bounds



(a) Greyscale image.

(b) Intensity histogram.

(c) Colour image.

(d) Colour histogram.

Figure 2.2.1: An image with its associated intensity and colour histograms.

of the histogram respectively. Generally, one chooses n and then partitions the space over the maximum and minimum values in the data, but it can be advantageous to keep a fixed range for the data, for example binning over 0 to 255 for intensity histograms regardless of the image contents, so that one may compare histogram bin values directly. The formula for an equi-width histogram with a total of m data points is

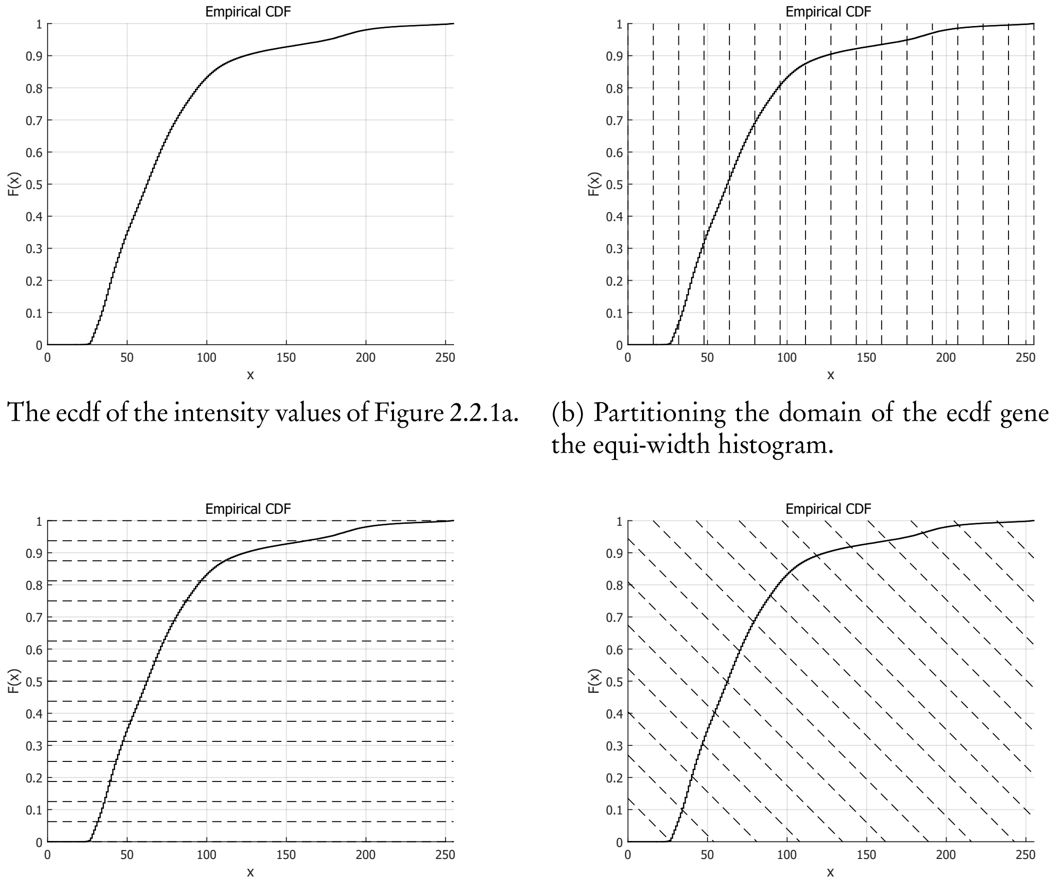
$$c_b = \sum_{j=1}^{m} \mathbb{1}(x_j), \tag{2.2.1}$$

where c_b is the count of data points in bin number $b \in \{1, \ldots, n\}$ and $\mathbb{1}(x_j)$ is an indicator function equalling 1 when bin b contains x_j and 0 otherwise. The bin b contains x_j if $\text{edge}_{left}(b) \leq x_j < \text{edge}_{right}(b)$. The last bin also includes the right edge, such that it contains x_j if $\text{edge}_{left}(b) \leq x_j \leq \text{edge}_{right}(b)$. Histograms are often normalised to have unit area, which corresponds to a total probability of 1. This removes the dependence on sample size and allows for different-sized sample distributions to be compared, resulting in a relative frequency histogram. The frequency f_b of bin b is calculated by

$$f_b = \frac{c_b}{\sum_{i=1}^{n} c_b}. \tag{2.2.2}$$

Piatetsky-Shapiro and Connell [36] studied the histogram as applied to database query optimisation, and showed that according to the maximum estimation error the key parameter to control in generating histograms is the depth, or height, of bins and not the width. They introduced the equi-depth histogram which takes the data and partitions it into n equal depth *distribution steps*. The process sorts the data using the natural ordering of the data domain and splits them into equal-sized steps. The effect of this is to create narrower bins in high density regions, and to have wider bins at regions of lower density. This makes intuitive sense if you are looking to characterise the "crux" of the data as you are paying more attention to more of the data. This, however, has the adverse effect of smoothing out outliers. This sort and step method works to make the bins exactly equi-depth, but what it fails to account for is the values sitting on the bin edge. These values would either get split across the bins depending on what index the step happens at, or if all of a bin's values were on the bin edge it would produce a zero-width bin. Scott [41] suggests taking instead a percentile mesh on the data, which does not create perfectly equi-depth bins, but negates the edge effects encountered from doing distribution steps.

The equi-width and equi-depth histograms both have minor shortcomings. The equi-width histogram does not handle spikes well, and has greater estimation error than the equi-depth histogram. However, it performs better at characterising outliers. By contrast, the equi-depth histogram gives greater detail to high-density regions but sacrifices resolution in the low-density regions. Denby and Mallows [13] propose a compromise between the two methods, which they term the diagonal histogram (d-hist). They use the empirical cumulative distribution function (ecdf) of the data to create their histograms, noting that equally-spaced partitions of the ecdf domain equates to creating an equi-width histogram, and equally-spaced partitions of the ecdf range equates to the equi-depth histogram. These partitions are illustrated in Figures 2.2.2b and 2.2.2c. From this perspective, the shortcomings of each can be explained: at high-density values the ecdf climbs sharply, which is lumped together by the equi-width histogram and better partitioned by the equi-depth histogram, whereas regions of low density will result in a flatter ecdf, which is better captured by the equi-width histogram and lumped together by the equi-depth histogram. Their proposition is to take equally-spaced diagonal partitions of the ecdf as in Figure 2.2.2d, which will make a



(a) The ecdf of the intensity values of Figure 2.2.1a.

(b) Partitioning the domain of the ecdf generates the equi-width histogram.

(c) Partitioning the range of the ecdf generates the equi-depth histogram.

(d) Diagonally partitioning the ecdf generates the diagonal histogram.

Figure 2.2.2: The way in which the empirical cumulative distribution function (ecdf) is partitioned determines the type of histogram generated.

compromise between capturing high-density regions, thereby capturing spikes, as well as capturing low-density regions, thereby capturing the outliers.

2.3 Histogram Distances

Histograms, being estimates of data distributions, have similarity measures (referred to as "distances") which can be calculated between them, much like their continuous counterparts. The distance, in effect, tells one how dissimilar two data distributions are, based on the estimates represented by the histograms. There are two main types of histogram distances: 1) the bin-to-bin type distances and 2) the cross-bin type distances. A comprehensive survey of distance measures is provided in [11]. For the following, let $P, Q \in \mathbb{R}^n$ be two vectors representing histograms with n bins.

2.3.1 Bin-to-bin distances

A bin-to-bin type distance computes the distance between the histograms in a bin-wise fashion. This is usually fast to compute, but can suffer from binning effects and histogram shifting. Binning effects occur when related values are incorrectly split into separate bins due to the position of the bin edge. Histogram shifting occurs when similar histograms are offset from one another, perhaps due to global effects such as lighting or shadows. This shift means incorrect bins are being compared when using bin-to-bin distances, as in Figure 2.3.1. The most common bin-to-bin distances are of the Minkowsky-form based on the L_p norm:

$$L_p(P, Q) = \left(\sum_i |P_i - Q_i|^p \right)^{\frac{1}{p}} \tag{2.3.1}$$

which is a metric for $p \geq 1$. From this family we get the most widely used distances, the L_1 or Manhattan distance, and the L_2 or Euclidean distance. A related distance that can handle histograms of different sample sizes is called the histogram intersection distance:

$$\mathrm{HI}(P, Q) = 1 - \frac{\sum_i \min(P_i, Q_i)}{\sum_i Q_i} \tag{2.3.2}$$

which is equivalent to the L_1 distance when both histograms are normalised.

The Kullback-Leibler (KL) divergence is an information-theoretic distance which measures how inefficient on average it would be to code samples from P when using a code based on Q:

$$\mathrm{KL}(P, Q) = \sum_i P_i \log \left(\frac{P_i}{Q_i} \right). \tag{2.3.3}$$

The KL divergence is non-symmetric, sensitive to histogram binning and equals infinity when $P \neq 0$ and $Q = 0$. To overcome some of these problems the Jenson-Shannon (JS) divergence was developed:

$$\mathrm{JS}(P, Q) = \mathrm{KL}(P, M) + \mathrm{KL}(Q, M) \tag{2.3.4}$$

where

$$M = \frac{P + Q}{2}.$$

From statistics we get the popular and effective χ^2 distance:

$$\chi^2 = \sum_i \frac{(P_i - M_i)^2}{M_i} \tag{2.3.5}$$

which measures how unlikely it is that one distribution was drawn from the population represented by the other, with M being the same as above.

Another statistical distance, the Bhattacharyya distance, started as a geometric similarity measuring the angle between two multinomial populations in a k-dimensional space, and it was shown in [3] that the Bhattacharyya coefficient could be derived to measure the similarity between two normalised histograms:

$$\text{BH}(P, Q) = \sum_i \sqrt{P_i Q_i}. \tag{2.3.6}$$

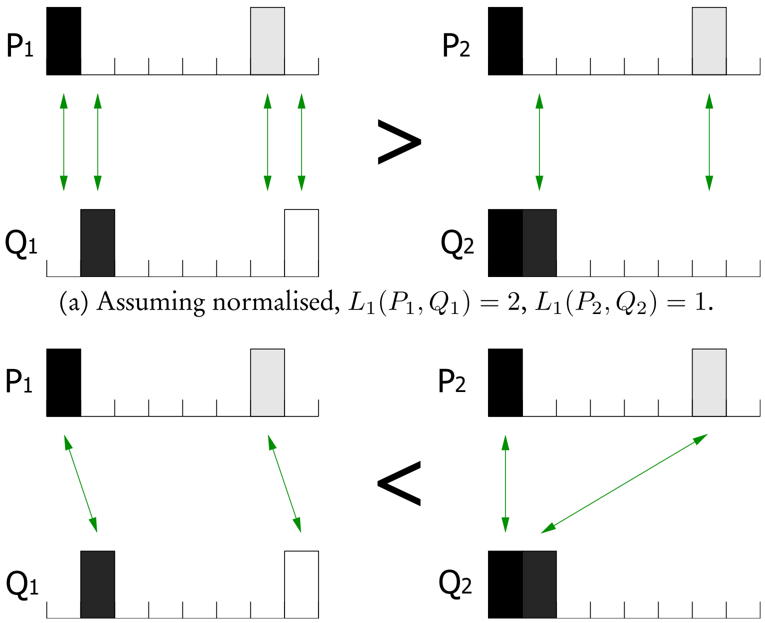
It is shown in [3] that the Bhattacharyya coefficient embeds the distance into a constant error space, which ensures that the minimum distance between two points is always a straight line between them. This property negates the need to evaluate the minimum of a curved path integral for Poisson like errors which one needs to do with the χ^2 distance. It is also shown that the Bhattacharyya distance approximates the χ^2 distance for small distances as well as avoids the singularity problem of χ^2 when comparing empty bins.

2.3.2 Cross-bin distances

Rubner [38] provides an illuminating illustration of the shortcomings of bin-to-bin distances. Figure 2.3.1 [38] shows that for the simple example of a shifted histogram the computed bin-to-bin distances do not match perception. This is due to the fact that the distances assume that the histograms are perfectly aligned. This makes them very sensitive to contrast and lighting changes which introduce shifts in the histogram. The bin-to-bin distances are also unduly affected by binning effects, which then greatly affects the distance produced.

Cross-bin distances attempt to overcome these shortcomings by taking into account the intra-bin values, which allows for a more robust comparison amongst histograms. The quadratic-form (QF) distance enlists a bin-similarity matrix $A \in \mathbb{R}^{n\times n}$, with a_{ij} being a value indicating how similar bin i is to bin j. The QF distance is then defined as

$$\text{QF}(P, Q) = \sqrt{(P - Q)^T A (P - Q)}. \tag{2.3.7}$$



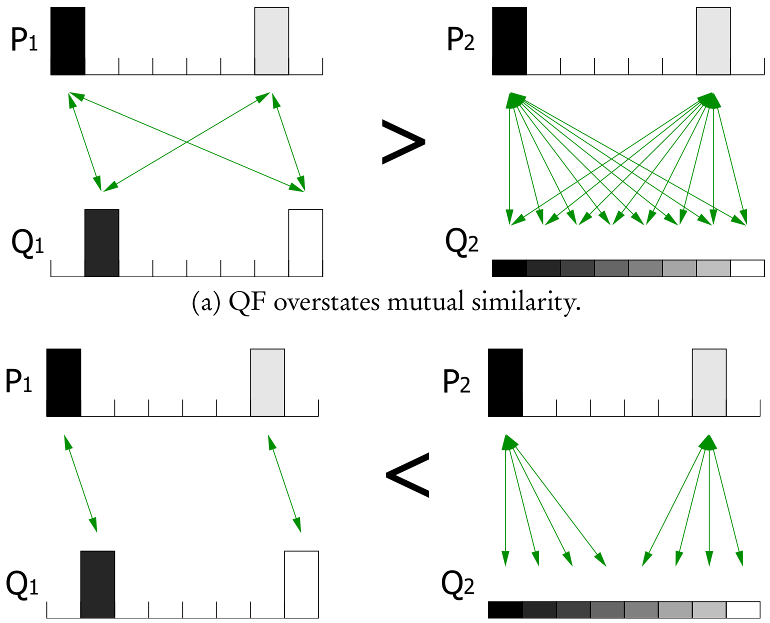
(a) Assuming normalised, $L_1(P_1, Q_1) = 2$, $L_1(P_2, Q_2) = 1$.

(b) Distance should be based on correspondence between bins.

Figure 2.3.1: An example of where the L_1 bin-to-bin distance does not match perception. Source: [38].

If A is chosen as the inverse of the covariance matrix, the distance reduces to the Mahalanobis distance. A common choice of A is to use the L_1 or L_2 distance between bins. The QF essentially maps each bin in one histogram with every bin in the other histogram, which overstates the mutual similarity for histograms without a pronounced mode. This produces distances which do not align with perception, as can be seen in Figure 2.3.2.

The earth mover's distance (EMD) [38] represents the distance between histograms as a transportation problem, also known as Monge-Kantorovich amongst others. It is a measure of how difficult it is to transform the smaller histogram into part of the bigger histogram, while taking into account the distances between bins. The sizes here refer to the number of samples in each histogram. It measures this difficulty of transformation by seeing the smaller histogram as piles of dirt in a space endowed with a ground distance, which is defined as the amount of work to move a unit of dirt from one location to another, and sees the bigger histogram as holes in that same space. The EMD is then the minimum amount of work to move all of the dirt into the holes.



(a) QF overstates mutual similarity.

(b) Distance should be based on correspondence between bins.

Figure 2.3.2: An example of where the QF distance does not match perception. Source: [38].

Formally:

$$\text{EMD}(P,Q) = \min_{\{F_{ij}\}} \frac{\sum_{i,j} F_{ij} D_{ij}}{\min(\sum_i P_i, \sum_j Q_j)} \quad \text{s.t.} \tag{2.3.8}$$

$$\begin{aligned} \sum_j F_{ij} &\leq P_i, \\ \sum_i F_{ij} &\leq Q_j, \\ \sum_{i,j} F_{ij} &= \min(\sum_i P_i, \sum_j Q_j), \text{ and} \\ F_{ij} &\geq 0. \end{aligned}$$

The optimal flow F_{ij} from P_i to Q_j is computed using the linear programming algorithm, the transportation simplex. Most research takes the ground distance D_{ij} between bin i and bin j to be the L_1 distance due to its attractive computational properties, but in this dissertation we will be using the L_2 distance to make use of the perceptual uniformity of colour in the CIELAB space, as explained earlier in Section §2.1.

The diffusion distance (DD) [32] is similar to the EMD, but instead of a transportation problem, it models the difference between histograms as a temperature field and relates the distance to an integrated norm of the diffusion process. For computational efficiency a Gaussian pyramid is used to discretise the continuous diffusion process, and the diffusion distance is then defined as the sum of norms over all pyramid layers. If $\mathbf{x} \in \mathbb{R}^n$ represents a point in the underlying data space, then the diffusion distance is defined as

$$\mathrm{DD}(P, Q) = \sum_{l=0}^{L} k(|d_l(\mathbf{x})|) \tag{2.3.9}$$

where

$$\begin{aligned} d_0(\mathbf{x}) &= P(\mathbf{x}) - Q(\mathbf{x}), \text{ and} \\ d_l(\mathbf{x}) &= [d_{l-1}(\mathbf{x}) * \phi(\mathbf{x}, \sigma)] \downarrow_2 \end{aligned}$$

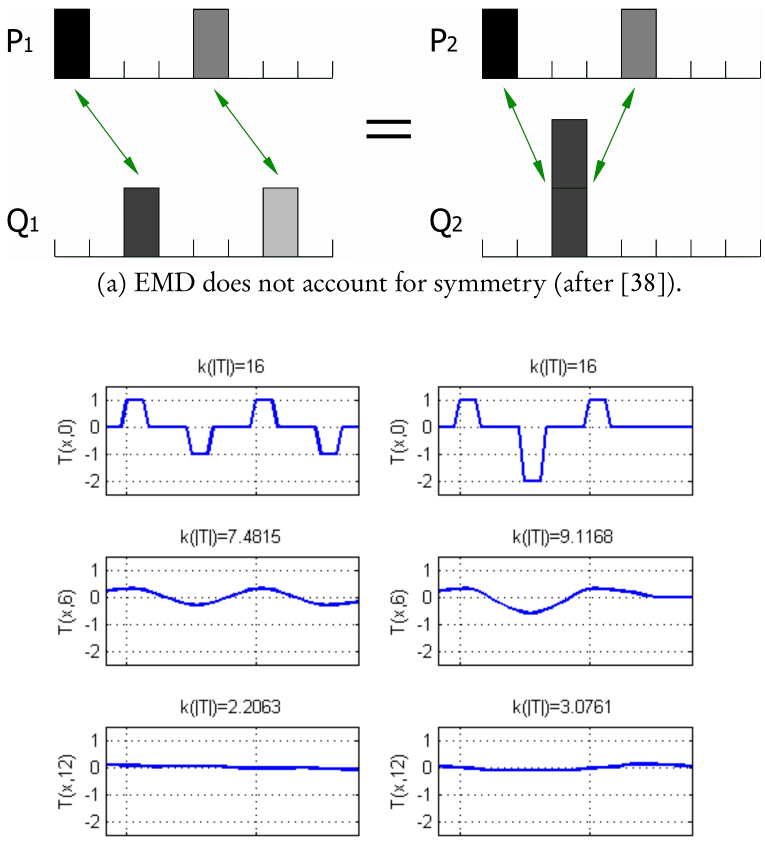
are different layers of the pyramid, with $\downarrow_2$ denoting half-sized downsampling, L being the number of layers, and $\phi(\cdot)$ a Gaussian filter with standard deviation σ. The authors note that as long as $k(\cdot)$ is a metric, $\mathrm{DD}(P, Q)$ also forms a metric on histograms. They choose to use L_1, reducing Eq. 2.3.9 to

$$\mathrm{DD}(P, Q) = \sum_{l=0}^{L} |d_l(\mathbf{x})|. \tag{2.3.10}$$

Using the example from [32] in the style of [38] it can be seen in Figure 2.3.3 that what would produce equal distances using the EMD is slightly better handled perceptually by the diffusion distance. The difference is marginal, but thanks to the downsampling and reuse of the Gaussian kernel, it produces results with a computational complexity of $\mathrm{O}(n)$.

2.4 Saliency

In understanding vision, Gestalt psychologists believe that the whole precedes the part, that we register unitary objects and only later, if necessary, do we analyse these objects further into components or properties. A turning point came when Treisman and Gelade introduced the feature-integration theory of attention [44], which states that features are registered early, automatically, and in parallel across the whole visual field, while objects are identified separately and only at a later stage, requiring focused attention. The physiological evidence [49] of specialised cells for features such as orientation and motion, coupled with numerous behavioural experiments, gives great weight to the



(a) EMD does not account for symmetry (after [38]).

(b) Diffusion distance captures the symmetry (after [32]).

Figure 2.3.3: An example of where EMD does not match perception.

validity of this theory.

In a world as complex as this it makes intuitive sense to have levels of processing. It would be prohibitively expensive and computationally taxing to maintain a complete representation of the entire world at every instant. Yet we still need to be able to identify potential predators, prey, mates, and complete basic survival tasks. One of nature's solutions to this challenge is the attentional mechanism found in humans and many animals. To quote William James [24]:

> "Every one knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought... It implies withdrawal from some things in order to deal effectively with others."

In other words, attention is the focus on one object or a small area of interest to the exclusion of the rest. This mechanism provides us with enough information to navigate our world while allowing us to allocate our finite brain's limited energy and processing power most effectively. For this system to be successful it requires a mechanism for directing the "spotlight" of attention, which in the case of vision is called visual saliency.

Visual saliency is a measure of how much a certain stimulus in the visual field stands out from its surrounds, or in other words how much it "pops out". A saliency map is a topographical map indicating the saliency of a location in the visual field, formed by combining elements of individual feature maps such as colour, edges, orientation and motion. Koch and Ullman proposed a theoretical foundation for the saliency map and how it might be implemented in the primate brain [27].

The first computational implementation of a saliency map was developed by Niebur and Koch [34] and further extended by Itti et al. [23]. Their model consisted of an input image, processed into a Gaussian pyramid [2], from which they compute intensity, colour, orientation and temporal change features. These features are linearly combined into a saliency map, having a greater weight for the temporal change feature based on perceptual observations made by the authors. They select the most salient location in the saliency map by means of a winner-take-all process. Interestingly, they also implement an inhibitory return signal, reasoning that the saliency map does not just find the most salient region and then stop, but rather it finds the most salient region, allows processing of it, finds the next most salient, allows processing of it, and so on. To allow for this, the inhibitory signal coming from the winner-take-all process applies a transient Mexican hat, or difference of Gaussians, at the location where the signal originated. This has the dual effect of inhibiting the most salient region but also of slightly raising the saliency of regions nearby, which might prevent the attention jumping too rapidly or drastically around the scene. For the interested reader, Itti and Koch [22] provide a more thorough review of the biologically plausible saliency maps and their computational counterparts.

Not all saliency maps are biologically based. Studies of the statistics of natural images show that there is an invariance to scale in natural images [39]. The property is known as the $1/f$ law and states that the amplitude $A(f)$ of the averaged Fourier spectrum of an ensemble of natural images obeys a distribution $\mathrm{E}\{A(f)\} \propto 1/f$. Hou and Zhang [18] show that the analysis of 2277 natural images revealed local linearity in the log spectrum of the images, with each image containing a similar trend with some statistical singularities. They reason that if there is a similarity in the log spectra across a wide variety of images, the information that deviates from these smooth curves is what should be attended to. They therefore apply a local averaged filter to the log spectrum to generate the smooth trend curve, and compute the spectral residual of the image as the difference between the local average spectrum and the image spectrum. The authors claim that the spectral residual contains the innovation of the image, and it is this innovation which is defined as being most salient. Due to this method being based on the global Fourier transform, which works directly on image intensity values, the results are not always aligned with what is perceptually salient.

Li et al. [31] explored this concept further and discovered that the spectral residual is of little significance, and show that by replacing it with random white noise with the same average value and maximum as the spectral residual they are able to achieve almost the same saliency map. They deduced that the spectral residual, which can be approximated by a horizontal plane, actually acts a high pass filter on the image. The amplitude spectrum of natural images always has higher amplitudes at lower frequencies, so when the amplitude spectrum is replaced by a horizontal plane it is, in effect, treating all frequencies equal. By virtue of this, the lower frequencies are suppressed and the higher frequencies are enhanced. This is almost equivalent to a gradient enhancement operation, which is why it discovers small salient objects but will only highlight the edges of larger objects and of textured regions in an image. They then turn the saliency identification on its head, and choose to search for nonsalient regions based on the fact that salient objects come in many shapes and forms, and can be spread across the image, whereas the backgrounds and nonsalient regions are generally repeating or uniform, which they then suppress to highlight the salient objects. The authors show that a repeated pattern in a signal corresponds to a sharp peak in the amplitude spectrum in the Fourier domain. Convolving the amplitude spectrum with a Gaussian kernel effectively suppresses the periodic background and nonsalient regions, leaving behind the salient objects which they highlight with some post-processing. The size of the smoothing kernel affects the size of the detectably salient region, so they introduce a scale-space representation and use the concept of entropy to select the appropriate scale. To include more features they replace the Fourier transform by the hypercomplex Fourier transform, using the opponent colour channels as the quaternion values.

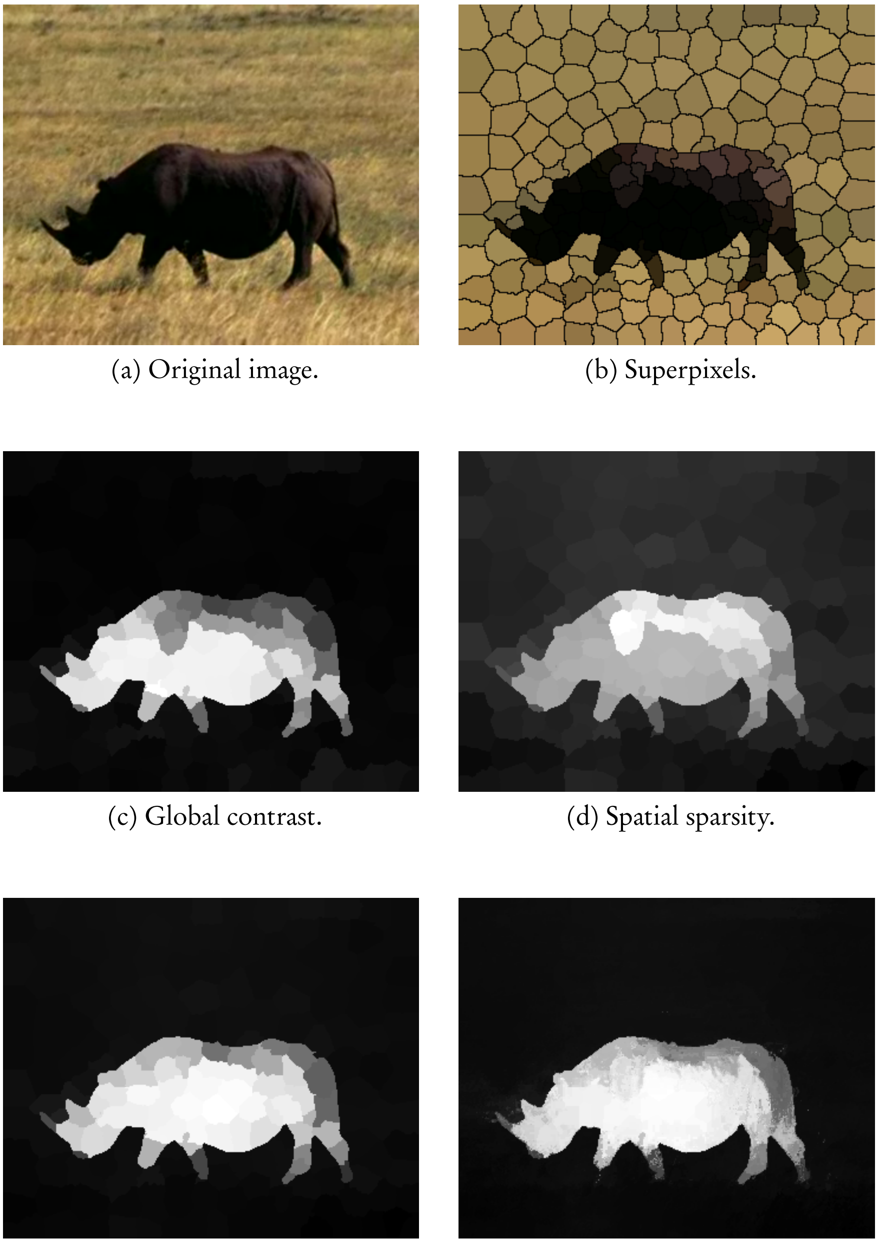
Bruce and Tsotsos [9] approach saliency from an information-theoretic standpoint, defining saliency in terms of the self-information of local patches of the image with respect to their surrounds. To the authors, saliency is synonymous with surprise, or the expected number of guesses it would take to predict the local patch based on its surroundings. To achieve this, a bank of filters is learned from a database of natural images using independent component analysis (ICA), forming a suitable basis of Gabor-like filters that correlate well with the V1 cortical cells found in the primate visual system. An estimate of the distribution of each basis coefficient is learned across the entire image via non-parametric density estimation. The probability of observing a local patch centered at any image location is then evaluated by independently considering the likelihood of each corresponding basis coefficient, with the product of all likelihoods yielding the joint likelihood of the entire set of basis coefficients. Shannon's measure of self-information is used to translate the joint likelihood into the resulting saliency map.

The rest of this dissertation focuses on the next model of saliency by Liu et al. [33], which, in a similar vein to Bruce and Tsotsos [9] above, uses global and local features to compute a saliency map. In particular, a statistical approach is taken, almost akin to outlier detection, whereby the global nature of the image is characterized by colour and motion histograms as features, and then compared via distance functions with smaller, homogeneous, edge-preserving regions called superpixels to generate the resulting saliency map. The model is made up of a colour saliency map and a motion saliency map, and is adaptively fused to generate a spatio-temporal saliency map. This dissertation looks at the specific effects histogram types and distances have in the calculation of colour saliency, which extend naturally to the temporal saliency map by virtue of using the same formulations. A brief review of the colour saliency map generation follows.

2.4.1 Colour Saliency

To aid in visualising the workings of this saliency model, Figure 2.4.1 shows the model output for an example image at various stages of the calculation. Each frame or image F_t (Figure 2.4.1a) is transformed into the CIELAB colour space, and is segmented into superpixels $sp_i (i = 1, \ldots, n)$, where n is the total number of superpixels (Figure 2.4.1b). Superpixels create an oversegmentation of an image, being generally homogeneous in colour and edge-preserving of the structures in the image [1, 46]. Working with superpixels is more meaningful in the context of an image, due to predefined blocks or circular regions destroying or ignoring the natural structure in an image.

Colour histograms are used as the features to determine colour saliency in an image. Each of the frame's CIELAB channels is uniformly quantised into q_b bins, generating a colour quantisation table CQ with $q_C = q_b \times q_b \times q_b$ bins. The authors set q_b to 16, claiming that it is sufficient for quantising colour images. Colour histograms, being 3-dimensional, have greater memory requirements than their 2-dimensional counterparts. However, by limiting the number of bins per histogram to 4096 and using superpixels to segment the image into a small number of regions, an upper limit on the memory requirement is induced. Using CQ, the global or frame-level colour histogram CH_0 is calculated using the entire frame's pixels, and normalised such that



(a) Original image.

(b) Superpixels.

(c) Global contrast.

(d) Spatial sparsity.

(e) Superpixel-level spatial saliency.

(f) Pixel-level spatial saliency.

Figure 2.4.1: The process for generating spatial saliency based on colour.

$\sum_{k=1}^{q_C} \mathrm{CH}_0(k) = 1$. The quantised colour for each bin, $qc(k)$, is calculated as the mean colour of all pixels that fall into bin k. Local or superpixel-level histograms, CH_i $(i = 1, \ldots, n)$, are then calculated and normalised such that $\forall sp : \sum_{k=1}^{q_C} \mathrm{CH}_i(k) = 1$.

The authors make two assumptions to generate their colour saliency map: 1) salient regions generally show contrast with the surrounding background regions, and 2) salient object colours are generally more sparsely distributed over the scene than background colours. They quantify the first assumption as the global contrast in the frame, which is defined by comparing each superpixel-level colour histogram with the frame-level histogram

$$S_{\mathrm{GC}}(sp_i) = \sum_{j=1}^{q_C} \left[\mathrm{CH}_i(j) \sum_{k=1}^{q_C} \|qc(j) - qc(k)\|_2 . \mathrm{CH}_0(k) \right] \tag{2.4.1}$$

where $\|\cdot\|_2$ is the L_2 norm. This states that the global contrast for a superpixel in relation to the frame is calculated as a sum of occurrence-weighted distances between the quantised colours present in the superpixel and the frame. Figure 2.4.1c shows the global contrast saliency value for each superpixel.

To quantify the second assumption of the colours of salient objects being more sparsely distributed, they define the spatial sparsity measure. To compute the spatial sparsity, each superpixel is compared with every other superpixel to create an intra-frame similarity value

$$\lambda_{intra}(sp_i, sp_j) = \sum_{k=1}^{q_C} \sqrt{\mathrm{CH}_i(k).\mathrm{CH}_j(k)} \,.\, \left[1 - \frac{\|\mu_i - \mu_j\|_2}{d} \right], \tag{2.4.2}$$

where μ_i and μ_j are the centroids of sp_i and sp_j respectively, and d is the diagonal length of the frame. The first term is the Bhattacharyya coefficient as described in Eq. 2.3.6, which measures the similarity between the two superpixel colour histograms, and the second term is a distance weighting function. The equation will evaluate higher for superpixels with more similar colour distributions to one another, and which are spatially closer to one another. Now for each superpixel the spatial spread of its colour distribution is calculated by

$$\mathrm{SD}(sp_i) = \frac{\sum_{j=1}^{n} \lambda_{intra}(sp_i, sp_j).\mathrm{D}(sp_j)}{\sum_{j=1}^{n} \lambda_{intra}(sp_i, sp_j)} \tag{2.4.3}$$

where $\mathrm{D}(sp_j)$ is the Euclidean distance from μ_j to the center of the frame. This, in effect, uses the center of the frame as a reference point to calculate the dispersion of the current superpixel's colour distribution. The spatial sparsity measure is then defined by an inverse normalisation of the above spread measure for each superpixel:

$$\mathrm{S_{SS}}(sp_i) = \frac{\max\left[\mathrm{SD}(sp)\right] - \mathrm{SD}(sp_i)}{\max\left[\mathrm{SD}(sp)\right] - \min\left[\mathrm{SD}(sp)\right]}. \tag{2.4.4}$$

The image in Figure 2.4.1d shows the spatial sparsity value for each superpixel.

The final spatial saliency value for each superpixel sp_i is then defined as a superpixel-wise multiplication of the global contrast with the spatial sparsity

$$\mathrm{S_S}(sp_i) = \mathrm{S_{GC}}(sp_i).\mathrm{S_{SS}}(sp_i), \tag{2.4.5}$$

and an example of the output at this stage is seen in Figure 2.4.1e.

To generate the pixel-level spatial saliency map, the pixel-level spatial saliency $\mathrm{S_S}(p_i)$ for each pixel p_i is defined as the sum of spatial saliency from the superpixel neighbourhood of the pixel, weighted by the pixel colour's probabilities in the corresponding superpixel-level colour histograms

$$\mathrm{S_S}(p_i) = \frac{\sum\limits_{sp_j \in \mathrm{N}(p_i)} \mathrm{S_S}(sp_j).\mathrm{CH}_j[\mathrm{bin}(p_i)]}{\sum\limits_{sp_j \in \mathrm{N}(p_i)} \mathrm{CH}_j[\mathrm{bin}(p_i)]} \tag{2.4.6}$$

where $\mathrm{N}(p_i)$ is the local neighbourhood of superpixels of p_i, including the superpixel containing p_i, and $\mathrm{bin}(p_i)$ denotes the entry number for the quantised colour of p_i in the colour quantisation table CQ. The superpixel neighbourhood is defined as all superpixels that make contact with the superpixel in question. The final saliency map with pixel-level saliency values is shown in Figure 2.4.1f.

The paper was implemented and validated against the results obtained for the DS1 dataset defined in their paper, and results are presented in Figure 2.4.2. The current implementation uses a different superpixel implementation from [46] and a newer optical flow method called edge-preserving patchmatch (EPPM) [4] which is much more efficient than the large displacement optical flow (LDOF) [8] used in the original paper, and also better preserves the edges of the flow boundaries. The current implementation marginally outperforms the original.

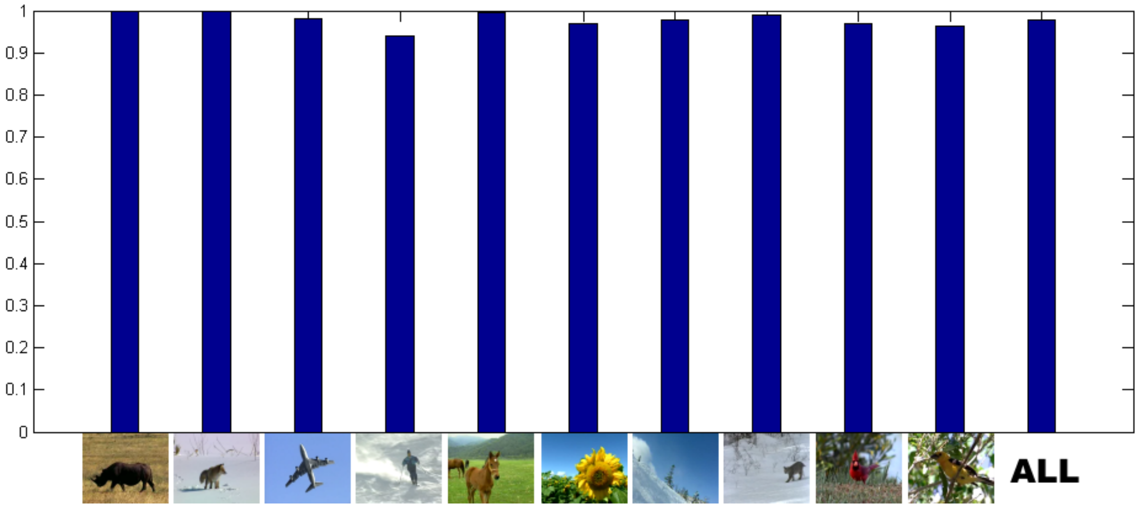


Figure 2.4.2: Saliency implementation AUC scores for dataset DS1 from [33].

2.5 Scoring Saliency

Saliency has its roots in the biological workings of the primate, and especially the human, visual system. It was shown by Hoffman and Subramaniam [17] and confirmed by Salvucci [40] that visual spatial attention and saccadic eye movements are related, finding that human subjects cannot move their eyes to one location and attend to a different one. This implies a tight coupling between visual attention and eye movements, and it is based on this fact that the most prevalent saliency scoring mechanisms are based on eye-tracking data of human subjects.

The most common form of saliency model test is to use the free-viewing task [28], which is accomplished by tracking the eye movements of human subjects using commercial grade eye-tracking systems while they freely view image or video databases. More recently, specific task-based viewing and object segmentations have also been used. Based on this eye-tracking data, a number of measures for how well a saliency map predicts or accounts for the spatial attention have been developed.

Eye-tracking data generally provides a set of (x_i, y_i) points of fixation per subject per image, but we know that the high quality foveal area covers approximately 2° of the visual field [20]. We can then create a map with 1 at each fixation point, and convolve it with a Gaussian the same size as the visual field, generating a heat map of where the subjects were looking. If we view both the saliency map S and the eye-tracking ground truth G as random variables, we can calculate the correlation coefficient (CC) as

$$\mathrm{CC(G, S)} = \frac{\sum_{x,y} \left(\mathrm{G}\left(x,y\right) - \mu_G\right) . \left(\mathrm{S}\left(x,y\right) - \mu_S\right)}{\sqrt{\sigma_G^2 . \sigma_S^2}} \tag{2.5.1}$$

where μ_G and μ_S are the means and σ_G^2 and σ_S^2 are the variances of the values in G and S respectively. This produces a single number in the range [-1, 1], where 0 indicates no correlation, 1 indicates perfect correlation and -1 indicates perfect anti-correlation between the two random variables.

The normalised scanpath saliency (NSS) [35] tests the correspondence of the human fixation points with the model-generated saliency maps. The model-generated saliency map is linearly normalised to have zero mean and unit variance:

$$\mathrm{S}_{\mathrm{norm}}(x, y) = \frac{\mathrm{S}(x, y) - \mu_S}{\sigma_S}. \tag{2.5.2}$$

Then, to account for inaccuracies in human fixation locations, the authors in [30] compute the NSS value for each fixation using a neighbourhood around the fixation:

$$\mathrm{NSS}(x_i, y_i) = \sum_{j \in \Omega} K_h(x_i - x_j, y_i - y_j) . \mathrm{S}_{\mathrm{norm}}(x_j, y_j) \tag{2.5.3}$$

where K is a kernel with bandwidth h and Ω is a neighbourhood. The NSS is then computed as the mean of NSS(x_i, y_i) for all fixations M of an observer:

$$\mathrm{NSS} = \frac{1}{M} \sum_{i=1}^{M} \mathrm{NSS}(x_i, y_i). \tag{2.5.4}$$

Due to the normalisation, positive values indicate a greater than chance correspondence of the human fixations with the saliency map, zero indicates no correspondence and negative values indicate anti-correspondence.

The most popular measure in the research is the area under the ROC curve (AUC) [43]. Receiver operating characteristic (ROC) is used to evaluate a binary classifier system by varying its discrimination threshold. The model-generated saliency map S is treated by a varying threshold on the saliency values, creating a binary fixation map for each level of the threshold. The human fixations are then used as the ground truth. The ROC curve is drawn as the false positive rate F_p (incorrectly labelling non-fixated locations as fixated) versus the true positive rate T_p (correctly labelling fixated locations as fixated), and the total area under the curve indicates how well the saliency model predicts human eye fixations. An AUC value of > 0.5 means the model is able to discriminate fixations from non-fixations greater than chance, 1 being perfect discrimination, and an AUC value of < 0.5 means the model performs worse than chance, with an AUC value of 0.5 meaning the model contains no discrimination power at all. As a variation of this, the human fixations are taken as the positive set, and some uniformly sampled points from the image are chosen as the negative set [7].

A problem has been identified in the literature, common to all saliency evaluation methods, which has been termed the center bias [45]. It has been observed through many experiments that subjects' fixation points are biased toward the center of static images as well as in videos. The issue arises in not being able to evaluate saliency models accurately, due to it being unknown as to whether the bias is induced by visually salient regions or by other contributing factors. One of the most prominent factors affecting the center bias is that of the photographer's bias, being that photographers generally place objects of interest towards the center of the frame. This is not bad in and of itself, due to photographers generally focusing on salient or interesting regions. Another factor, known as the viewing strategy, is when subjects reorient upon new stimuli with greater frequency toward the center of the frame, usually after repeated exposure to photographer-biased stimuli. This is also due to many datasets requiring subjects to fixate on the center of the screen prior to being shown a new stimulus.

The problem of center bias is illustrated by Zhang et al. [50] and Judd et al. [25] when they use a centered Gaussian blob as the saliency map and obtain much greater than chance AUC scores, even higher than some saliency models. In the recent review of saliency scoring methods, Borji et al. [7] show that the center bias and smoothing of the fixation points into a heatmap affect all scores previously mentioned. They propose the shuffled AUC (sAUC) [50] as a viable measure, whose only difference is instead of taking a uniform sampling from the image as the negative set, all fixations from all other images are used as the negative set. They show that the sAUC value for a centered Gaussian is close to 0.5, meaning that it manages to account for the center bias sufficiently.

2.5.1 Image Dataset and Viewing Tasks

The selection of a specific dataset can have a big influence on the saliency scores generated. The number of images and the diversity of categories they belong to can play an important role in determining accurately and objectively how well a saliency model is

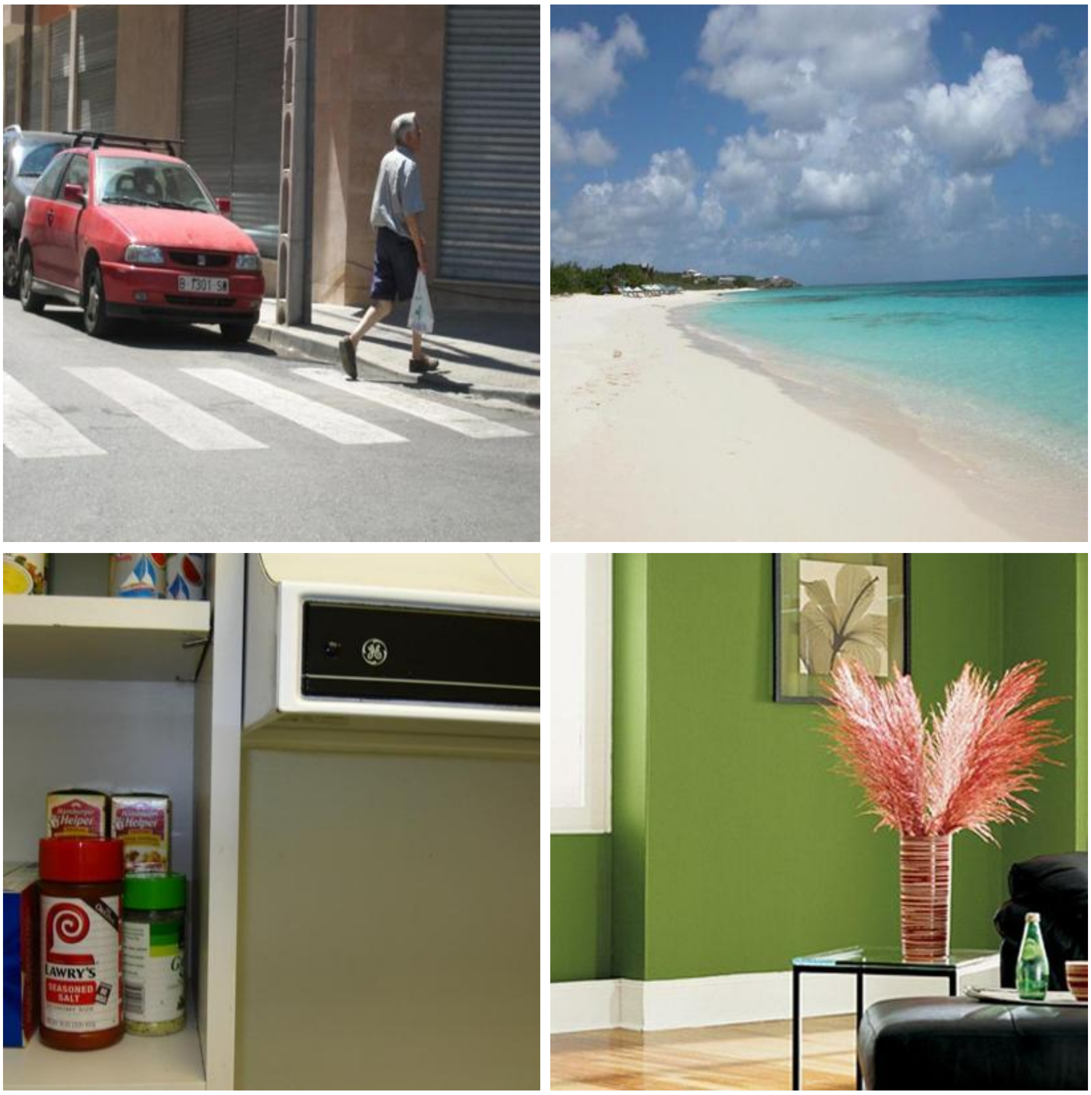


Figure 2.5.1: Example images from the dataset.

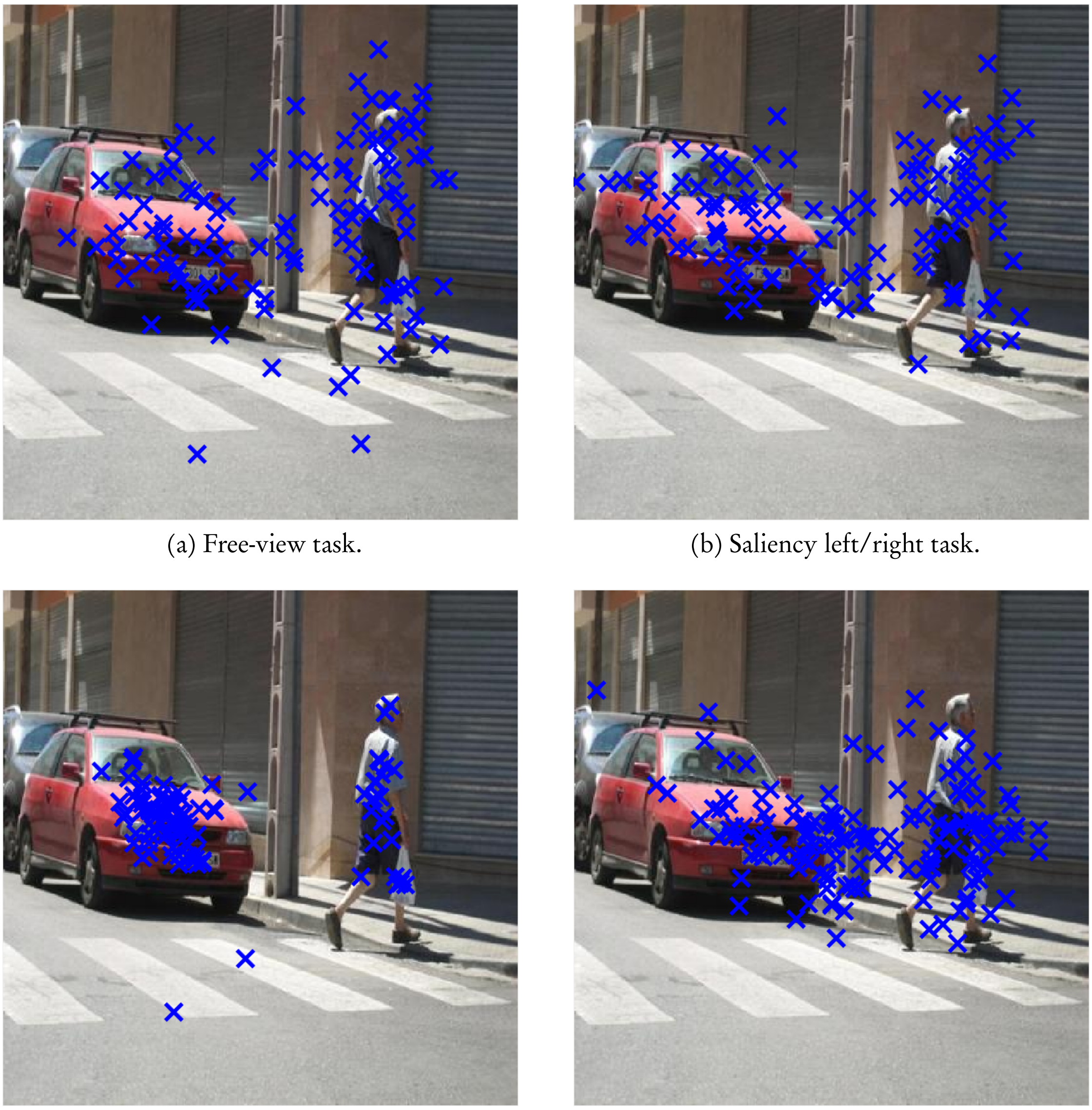
performing. The image dataset[3] selected for this dissertation was introduced recently in [28]. It consists of 800 photos of both indoor and outdoor scenes, either taken by the authors or obtained from existing datasets and online search engines. The images were specifically chosen to contain lateral (left/right) contextual information of a tangible object. Example images are shown in Figure 2.5.1.

Using this dataset, the authors attempt to uncover the relationship between bottom-up and top-down saliency with a host of carefully crafted experiments. What makes this dataset unique is that they provide eye-tracking data for three different tasks, as well as

[3] The dataset can be found at `https://labs.psych.ucsb.edu/eckstein/miguel/research_pages/saliencydata.html`.

mouse click location data for an explicit selection task. The images were displayed to participants so as to subtend 15° × 15° of visual angle, and were shown on a grey background. First was the typical free-viewing task with recorded eye movements. Second, participants were asked to decide whether the left or right half of the image was more salient while tracking their eyes. Third, participants were asked to explicitly select the object or region from the images that they considered to be most salient using mouse click selections. And fourth, participants were cued with object descriptions prior to viewing an image, and asked to report whether the object was present or not, which was missing 50% of the time, also while having their eyes tracked. The definition of "salient" given to the participants was something that stood out or caught their eye. They were given an example of a red flower among a field of white daisies when prompted for clarification. Examples of the recorded eye tracking data and explicit click selections are displayed in Figure 2.5.2.

An interesting finding of this paper was that saliency models were better able to predict the explicit saliency judgement tasks. One of the possible reasons given is that free-viewing is not without a top-down goal, but rather each individual would have an intrinsic goal or set of goals in the absence of an extrinsic one, which could vary across participants, and even for the same participant over many trials. When told explicitly to determine the salient regions or objects, it is thought the goals of the top-down and bottom-up systems have aligned.



(a) Free-view task.

(b) Saliency left/right task.

(c) Explicit-click task.

(d) Object proposal task.

Figure 2.5.2: Example eye-tracking and explicit click data displayed on an image from Figure 2.5.1.

Chapter 3
Experiment 1: The Effects of Histogram Construction Methods on Saliency Detection

As seen in Section §2.2, histograms come in many shapes and forms, and Section §2.4.1 shows that histograms can play a fundamental role in saliency detection. The literature that uses histograms rarely uses anything other than the most common technique for histogram construction, namely the equi-width histogram. Due to the known shortfalls of the equi-width histogram, such as being susceptible to binning effects, incorrectly characterising the data or not partitioning the space finely enough, the question of what impact the histogram construction method has on saliency detection arises. This chapter details the experiment conducted in order to answer this question. The aim of the experiment is clearly stated in Section §3.1, followed by the hypothesis in Section §3.2 and details of the tools utilised are provided in Section §3.3. The methods employed by the experiment are laid out in Section §3.4, with the results and conclusions following in Sections §3.5 and §3.6 respectively.

3.1 Aim

The aim of this experiment is to elucidate the role the histogram construction method plays in detecting colour saliency in images.

3.2 Hypothesis

A histogram is used for saliency detection primarily for two reasons, firstly to make the computation tractable by greatly reducing the computational requirements, and secondly to summarise the data so as to illuminate the underlying structure of the image to identify salient regions. The proposed hypothesis is that being better able to characterise the image by using variable bin-width construction methods will result in improved saliency detection.

3.3 Apparatus

The histogram construction methods under consideration will be the methods described in Section §2.2. As previously mentioned, histograms are used to summarise the data, which is also known as clustering. As a benchmark for how well histograms summarise the data, a k-means clustering algorithm will be used for comparison. The way k-means clustering works is to initialise k clusters by their means in the data space using a variety of methods, often by randomly placing them in the data space or using some form of sub-sampling of the data. Each data point is then classified by those k means by assigning it to the closest cluster mean. Each cluster mean is then recomputed based on the members in the cluster, and this process of reclassifying the data and recalculating the means is iterated until some specified convergence threshold is reached. This process can also be repeated multiple times with different initialisations and the best run's solution kept. It is a fairly simple approach to clustering, and is sensitive to outliers, but due to its more flexible structure is better able to capture the underlying structure in the image than the fixed nature of histograms. In this experiment k is chosen to be 256 as it provided a good tradeoff between performance and computation time, allowing comparison with the histogram-based methods.

The experiment was conducted on an Asus UX303LN laptop computer, containing an Intel i7-4510U 2.0GHz CPU, 12GB RAM and 256GB SSD HDD, running MATLAB R2013a on Microsoft Windows 8.1.

3.4 Method

3.4.1 Quantisation Error

One of the primary uses for histograms are for their summarisation ability. We can test how well they summarise an image's pixel data by computing the quantisation error when using each of the histogram construction methods to quantise our image dataset [10, 14]. This is equivalent to clustering the data and then using the cluster representative, in our case the mean colour, for each member's pixel colour and comparing it to the original image. This will give an indication of how well the histogram captures the underlying information in an image, and will allow us to make a comparison with the effect of the histogram construction method on detecting saliency.

Each image is quantised in both the RGB and CIELAB spaces by using each of the histogram construction methods detailed in Section §2.2, namely the equi-width, equi-depth and diagonal histograms ($\alpha = 5$ as per the authors' recommendation), as well as with the k-means algorithm ($k = 256$). The experiment is conducted in both the RGB and CIELAB spaces so as to determine whether the histogram construction methods are sensitive to the chosen colour space. The mean squared error (MSE) [14]

$$\text{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[\left(R_{ij} - \hat{R}_{ij}\right)^2 + \left(G_{ij} - \hat{G}_{ij}\right)^2 + \left(B_{ij} - \hat{B}_{ij}\right)^2\right] \qquad (3.4.1)$$

is then calculated for the RGB space, with (R, G, B) and $(\hat{R}, \hat{G}, \hat{B})$ being the RGB pixel values at position (i, j) for the original and quantised image of height M and width N respectively. The delta E (ΔE) metric [14]

$$\Delta\text{E} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{(L_{ij} - \hat{L}_{ij})^2 + (a_{ij} - \hat{a}_{ij})^2 + (b_{ij} - \hat{b}_{ij})^2} \qquad (3.4.2)$$

is calculated for the CIELAB space as the Euclidean distance between CIELAB pixel values (L, a, b) and $(\hat{L}, \hat{a}, \hat{b})$.

3.4.2 Histogram Construction Method

To determine the effect of the histogram construction method on saliency detection, the colour saliency calculation from Section §2.4.1 is computed using each of the construction methods as above to construct the global and local colour histograms, CH_0 and CH_i respectively. The resulting saliency maps are then scored using the metrics explained in Section §2.5 (CC, NSS, AUC, sAUC). This will highlight the differences between the various construction methods and their effects on saliency detection.

3.5 Results

3.5.1 Quantisation Error

Figure 3.5.1 shows an example image and the different quantised images with their L_1 errors. It can be seen in Figure 3.5.1b that the equi-width histogram is not able to

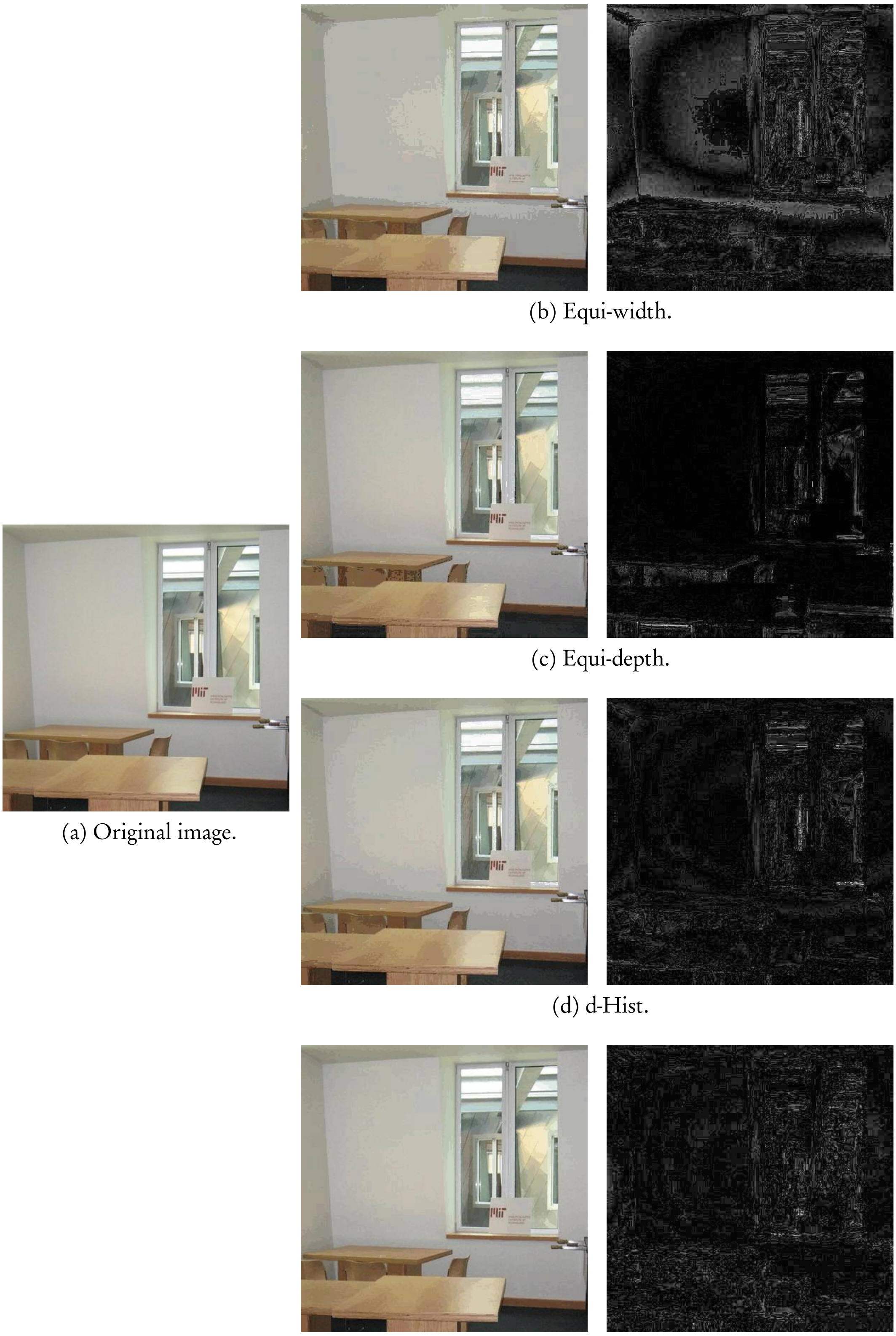
Table 3.5.1: Mean run times for computing the quantised image in seconds (error indicates one standard deviation).

Equi-width	**Equi-depth**	**d-Hist**	**k-Means**
0.168 ± 0.0158	0.280 ± 0.0279	0.336 ± 0.0259	53.7 ± 13.7

characterise the slight gradient of the wall colour due to its fixed bin widths, whereas the equi-depth histogram is able to provide more granularity to the densely-populated colours in the image. This means the error will be higher in the low-density colour regions, as can be seen by the brighter spots in Figure 3.5.1c, but due to this low density they have a small effect on the overall quantisation error. The d-hist method makes a compromise between the equi-width and the equi-depth histograms, as can be seen by the slightly higher error in the wall gradient, but then slightly less error in the windowsill's low-density colour features, as indicated in Figure 3.5.1d. The k-means method has a much smaller colour palette (256 versus 4096 colours), but due to its more flexible nature, the errors are spread fairly evenly across the colour space, and therefore across the image, as shown in Figure 3.5.1e. The results in Figure 3.5.2 indicate that the equi-depth histogram provides a smaller quantisation error in the RGB space, and a virtually on-par result with the d-hist method in the CIELAB space. It also matches the k-means method for quantisation error, and does so in a fraction of the time. The results also show an invariance to the selected colour space, with proportional quantisation errors between the colour spaces. Table 3.5.1 shows the average time for quantising the image running on an Intel i7-4510U 2.0GHz CPU, which shows just how much more efficient some histogram methods are.

3.5.2 Histogram Construction Method

Figure 3.5.3 shows some example saliency maps generated using the different histogram construction methods. Due to the equi-depth histogram assigning more bins to the higher-density regions, the lower-density regions are forced to spread out into the remaining bins. This tends to lump together the low-density or outlier regions, having an equalising and spreading effect on the saliency map as can be seen in the top row. The road and sidewalk are fairly homogenous in colour and so require more bins to characterise, leaving the sparser colours like the red sign and car shadows to be given equal weighting. Images with small salient regions and colours amongst large homogenous backgrounds will be most affected, with very busy and non-homogenous images starting to look like the equi-width based saliency. The bottom row shows the similarity



(b) Equi-width.

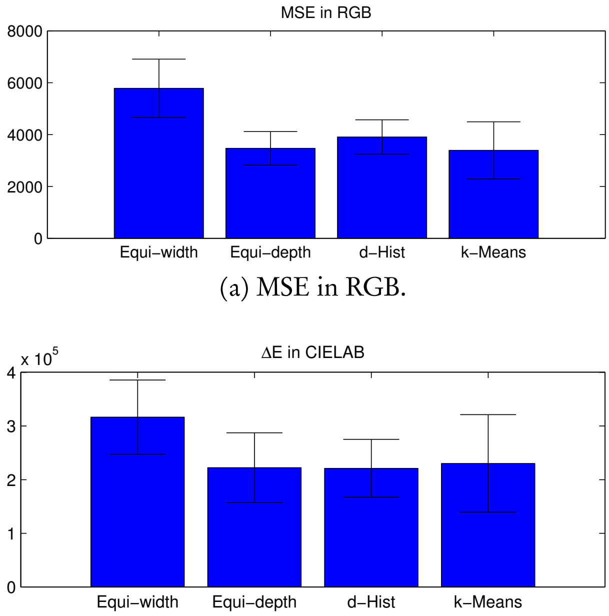
(c) Equi-depth.

(a) Original image.

(d) d-Hist.

(e) k-Means.

Figure 3.5.1: An example image with its quantised images and L_1 errors.



(a) MSE in RGB.

(b) ΔE in CIELAB.

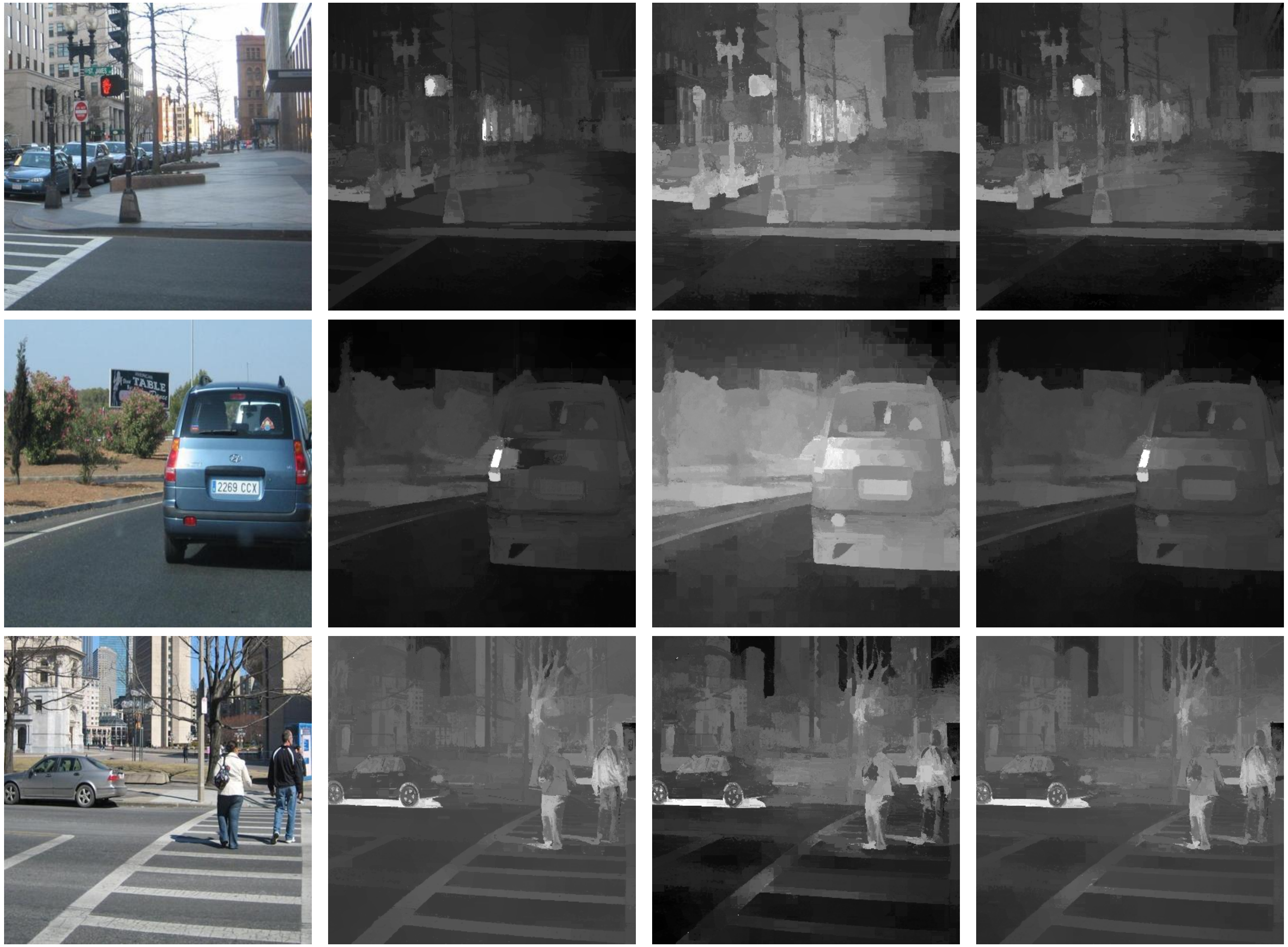
Figure 3.5.2: Quantisation error in RGB and CIELAB spaces using different histogram construction methods (error bars indicate one standard deviation).

of the saliency maps when the distinction between background and saliency becomes more difficult to identify. The d-hist is again a compromise between the equi-width and equi-depth histogram methods, a softer equalising that still retains some of the peaks of the equi-width saliency map.

Figure 3.5.4 shows the results of computing the aforementioned metrics on the image database for each task. One cannot help but notice the large variance of the NSS metric, especially for the explicit-click task (Figure 3.5.4c). To understand why this is so, Figure 3.5.5 shows the 3 highest and lowest NSS scoring saliency images overlaid with the participants' explicit-click locations along with their equi-width saliency maps. The high NSS scores are thanks to the saliency map being very localised, and most of the click positions landing on the high saliency locations. The low NSS scores highlight a feature of this saliency model which is not obvious from the outset, that being its built-in center bias. Recall Eq. 2.4.3, which is a weighted sum of distances to the center of the frame, weighted by intra-superpixel similarity and distance. This evaluates higher when there are more similar superpixels closer together and they are further from the center of the frame, and especially for both. The inverse normalisation in Eq. 2.4.4 then reverses this so that superpixels that were closer to the center, and not similar to any others will be ranked as more salient. This can clearly be seen in the low NSS scoring examples in Figures 3.5.5c and 3.5.5d, where the red blanket, pool and red couch all have much lower saliency, although they are salient in terms of global contrast, as seen in Figure 3.5.7, as well as for humans as pointed out by the human click locations.

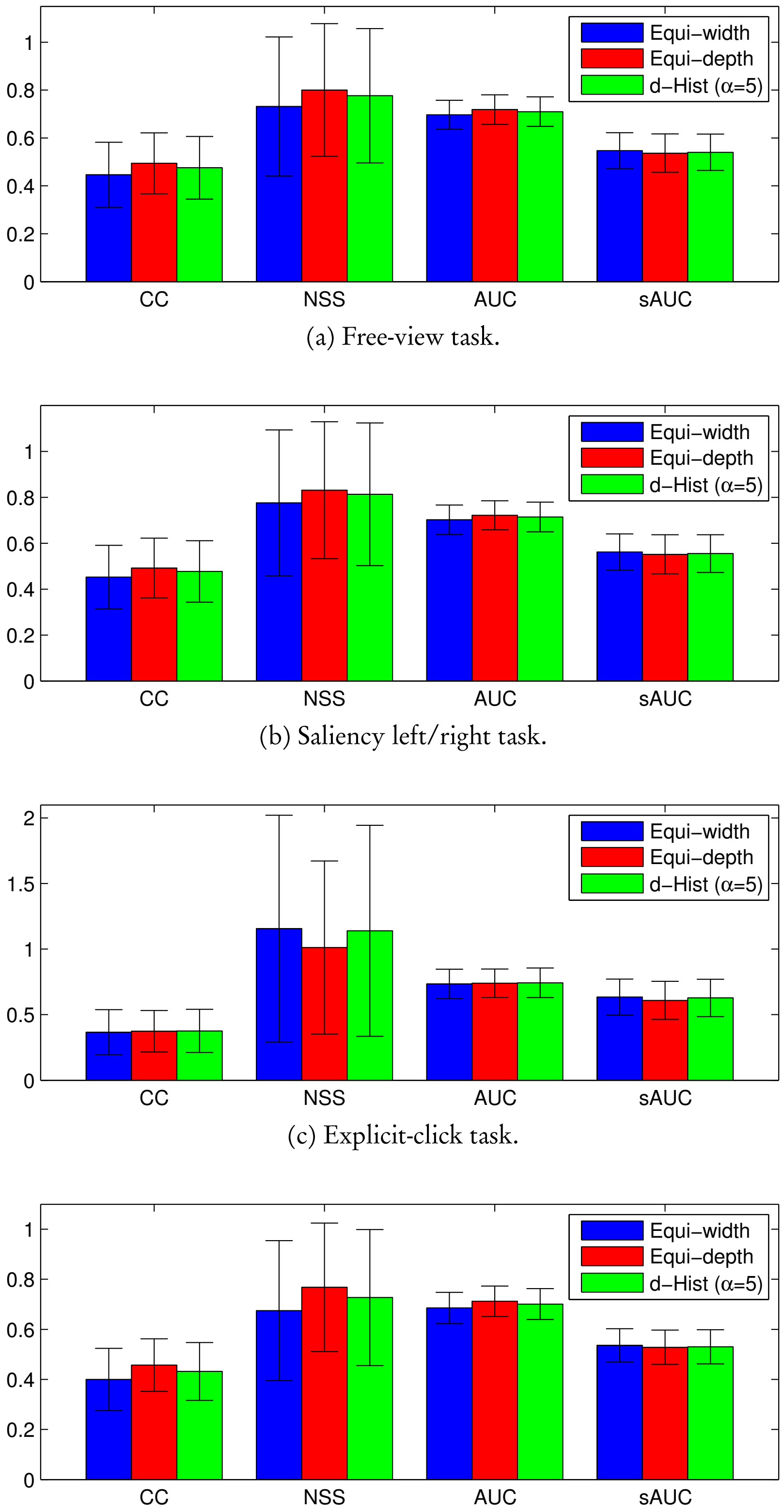
For all of the eye-tracked tasks, the equi-depth histogram construction method marginally improves all the metrics over the equi-width histogram, except for the sAUC where it remains practically constant. Figure 3.5.6 shows the top 3 and bottom 3 NSS scoring images overlaid with the participants' fixation locations with their equi-depth histogram constructed saliency maps. It can be seen here and in Figure 2.5.2 that the eye-tracking data is much more noisy than the explicit-click data, as well as having a heavy center bias. A possible reason for the marginal score increase might come from the equi-depth saliency maps being more "spread out" and not so peaky, which in turn maps better onto the more spread out eye-tracked fixation locations. The high-scoring maps correlate fairly well with the eye-tracking data, whereas with the low-scoring maps the eye-tracking data is either spread uniformly about the image, which might represent there not being anything universally salient in the image, or there are some forms of higher context in the image, such as people, faces or recognisable objects which naturally draw our attention [25], which colour alone is not able to characterise.

The equi-depth histogram causes a drop in NSS score for the explicit-click task (Figure 3.5.4c), which can be explained by the opposite reasoning as above. The click locations are much more accurate and grouped more tightly together, which means that if the saliency map has higher peaks at the click locations, it will result in higher NSS scores. This peakiness also means that when it misses, the resulting score will be much lower, as can be seen in the larger variance of the equi-width NSS score.



(a) Original. (b) Equi-width. (c) Equi-depth. (d) d-Hist.

Figure 3.5.3: Example images and their saliency maps computed with different histogram construction methods.



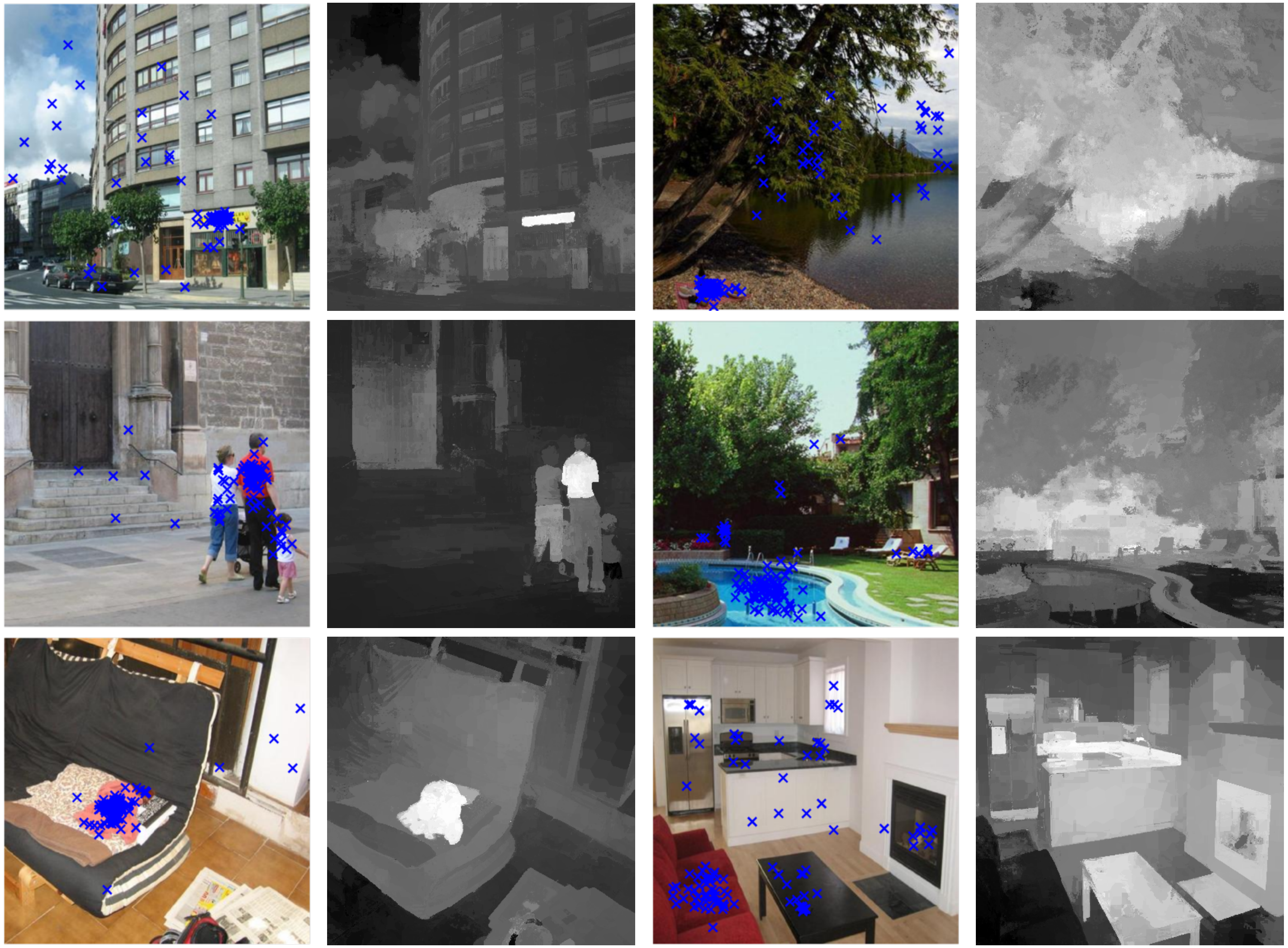
(a) Free-view task.

(b) Saliency left/right task.

(c) Explicit-click task.

(d) Object proposal task.

Figure 3.5.4: Results for using different histogram construction methods to detect colour saliency (error bars indicate one standard deviation).



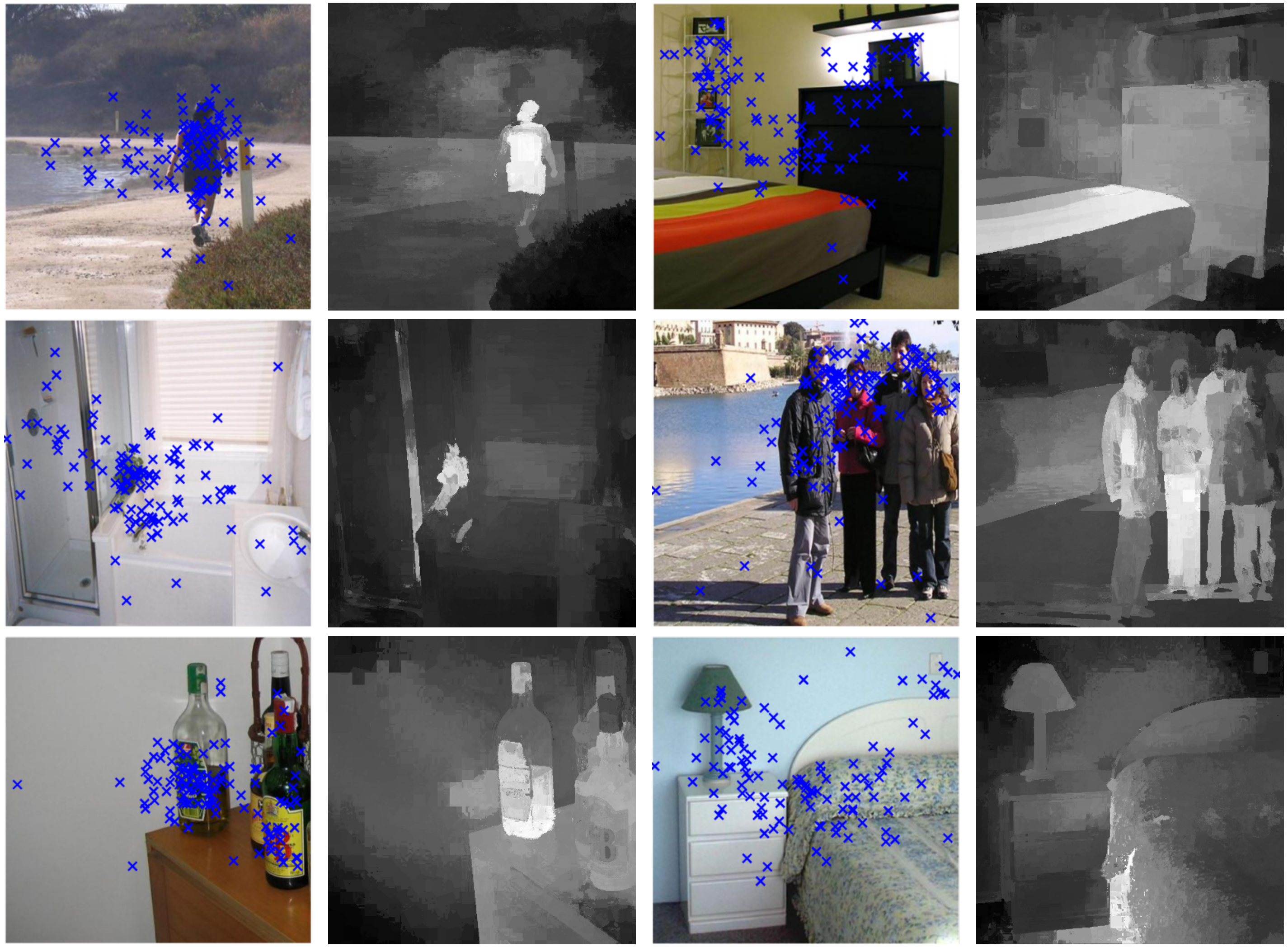
(a) High NSS with click locations overlaid.

(b) Equi-width.

(c) Low NSS with click locations overlaid.

(d) Equi-width.

Figure 3.5.5: The 3 highest and lowest NSS scores' images and their equi-width saliency maps for the explicit-click task.



(a) High NSS with click locations overlaid.

(b) Equi-depth.

(c) Low NSS with click locations overlaid.

(d) Equi-depth.

Figure 3.5.6: The 3 highest and lowest NSS scores' images and their equi-depth saliency maps for the free-viewing task.

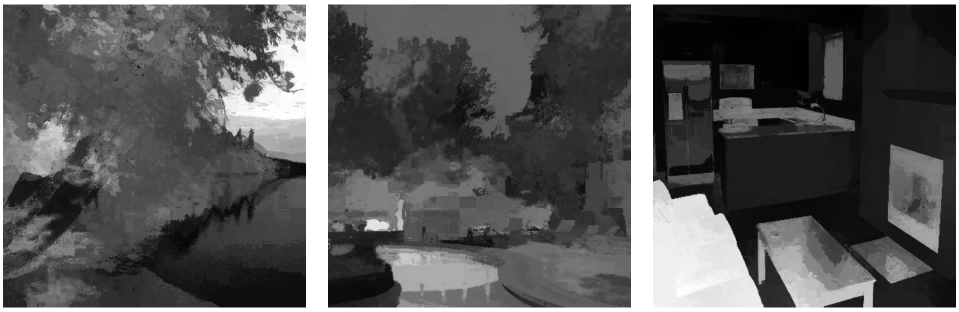


Figure 3.5.7: Global contrast only saliency maps of the images in Figure 3.5.5d.

3.6 Conclusion

The aim of this experiment was to investigate the role of the histogram construction method in saliency detection. It was shown that equi-depth histograms are better able to characterise the underlying structure of an image and produce less quantisation error than the more common equi-width histogram. It does so by providing more bins, and therefore more quantised colours, to the more densely populated colours of the image. This means the errors are larger for the low-density colours. They, however, have less effect on the quantisation error due to their low density.

It was also shown that for the eye-tracking tasks the equi-depth histogram was able to marginally improve the saliency scores. This was largely attributed to the equi-depth histogram assigning more bins to the high-density colours, forcing the low-density colours to clump together. This clumping equalises the saliency values of the low-density colours and causes the saliency map to become less peaky and more spread out, which correlates better with the noisy eye-tracking fixation data. The opposite holds true for the more accurate explicit-click task, due to the higher peaks leading to higher scores, albeit with greater variance due to lower scores when the clicks do not match the saliency map.

Chapter 4
Experiment 2: The Effects of Histogram Distances on Saliency Detection

As indicated in Section §2.3, there are many ways to compare histograms to one another. The colour saliency calculation in Section §2.4.1 uses two distance measures between histograms to calculate the saliency of a region in the image. The global contrast (Eq. 2.4.1) uses a type of cross-bin distance to calculate the distance between a superpixel's colour distribution and the frame's colour distribution, and the intra-frame similarity (Eq. 2.4.2) uses the Bhattacharyya coefficient (Eq. 2.3.6) to calculate the similarity between superpixel colour distributions. Due to the variety of distance methods, as well as their varying strengths and weaknesses, we would like to identify the effect the histogram distance measure has on computing the colour saliency in an image. This experiment is performed to measure this effect. The aim, hypothesis, apparatus and method are stated in Sections §4.1, §4.2, §4.3 and §4.4 respectively. This is followed by an explanation of how the results are presented in Section §4.5. The results are presented in Section §4.6 and the conclusion is drawn in Section §4.7.

4.1 Aim

The aim of this experiment is to determine what effect, if any, the histogram distance measure has on detecting saliency in an image.

4.2 Hypothesis

Bin-to-bin distances have the benefit of being simple to implement and can be computationally attractive to use. They, however, cannot handle shifts in histograms, which are easily produced by lighting and shading in an image, and are sensitive to binning effects, as illustrated in Section §2.3. Cross-bin distances, on the other hand, handle histogram shifts much better and do not suffer from binning effects, but are computationally expensive to run. The hypothesis of this experiment is that cross-bin distances will improve the saliency scores generated by negating the effects mentioned, but due to perceptually-similar regions being close distance-wise in the CIELAB space they will not improve the saliency score sufficiently to warrant the greatly increased computation time.

4.3 Apparatus

The experiment was conducted on an Asus UX303LN laptop computer, containing an Intel i7-4510U 2.0GHz CPU, 12GB RAM and 256GB SSD HDD, running MATLAB R2013a on Microsoft Windows 8.1.

4.4 Method

The original implementation uses two distance measures to compute the colour saliency, namely the global contrast (Eq. 2.4.1) and the spatial sparsity (Eq. 2.4.4). This experiment substitutes the distance measure in each of those equations and calculates the spatial saliency using Eq. 2.4.5. For the global contrast equation the distances computed are the original or pseudo-quadratic-form (PQF), the Jenson-Shannon divergence (JS) (Eq. 2.3.4), the earth mover's distance (EMD) (Eq. 2.3.8), and the diffusion distance (DD) (Eq. 2.3.10). The spatial sparsity, which is made up from the intra-frame similarity (Eq. 2.4.2), will be computed with the original or Bhattacharyya coefficient (BH) (Eq. 2.3.6), JS, EMD, and DD. When referring to which combination is used to produce the final spatial saliency map, the spatial saliency equation is modified to:

$$\mathrm{S_S}(\mathrm{GC_{dist}}, \mathrm{SS_{dist}}) = \mathrm{GC_{dist}}.\mathrm{SS_{dist}}, \tag{4.4.1}$$

where $\mathrm{GC_{dist}} \in \{\mathrm{PQF}, \mathrm{JS}, \mathrm{EMD}, \mathrm{DD}\}$ is the global contrast computed with the referenced distance and $\mathrm{SS_{dist}} \in \{\mathrm{BH}, \mathrm{JS}, \mathrm{EMD}, \mathrm{DD}\}$ is the spatial sparsity, which is computed using the intra-frame similarity with the referenced distance. This gives a total of 16 combinations of spatial saliency. For example, the original spatial saliency is referenced as $\mathrm{S_S}(\mathrm{PQF}, \mathrm{BH})$ and the spatial saliency using EMD for the global contrast and DD for spatial sparsity is referenced as $\mathrm{S_S}(\mathrm{EMD}, \mathrm{DD})$. All calculations for this experiment use the equi-width histogram as per the original paper [33].

4.5 Presentation of Results

The 4 combinations of global contrast and 4 combinations of spatial saliency create 16 saliency maps, with which 4 metrics (CC, NSS, AUC, sAUC) are calculated on the 4 tasks from the dataset detailed in Section §2.5.1, giving a total of $4 \times 4 \times 4 \times 4 = 256$ scores. To collate all this information the grid system in Table 4.6.1 is used.

At the finest level, the individual cells will represent the scores computed using combinations of distances for global contrast and spatial sparsity (the top magnified grid). These 4×4 grids make up the cells to a larger 4×4 grid which score the saliency maps using different scoring measures across different dataset tasks (the bottom grid).

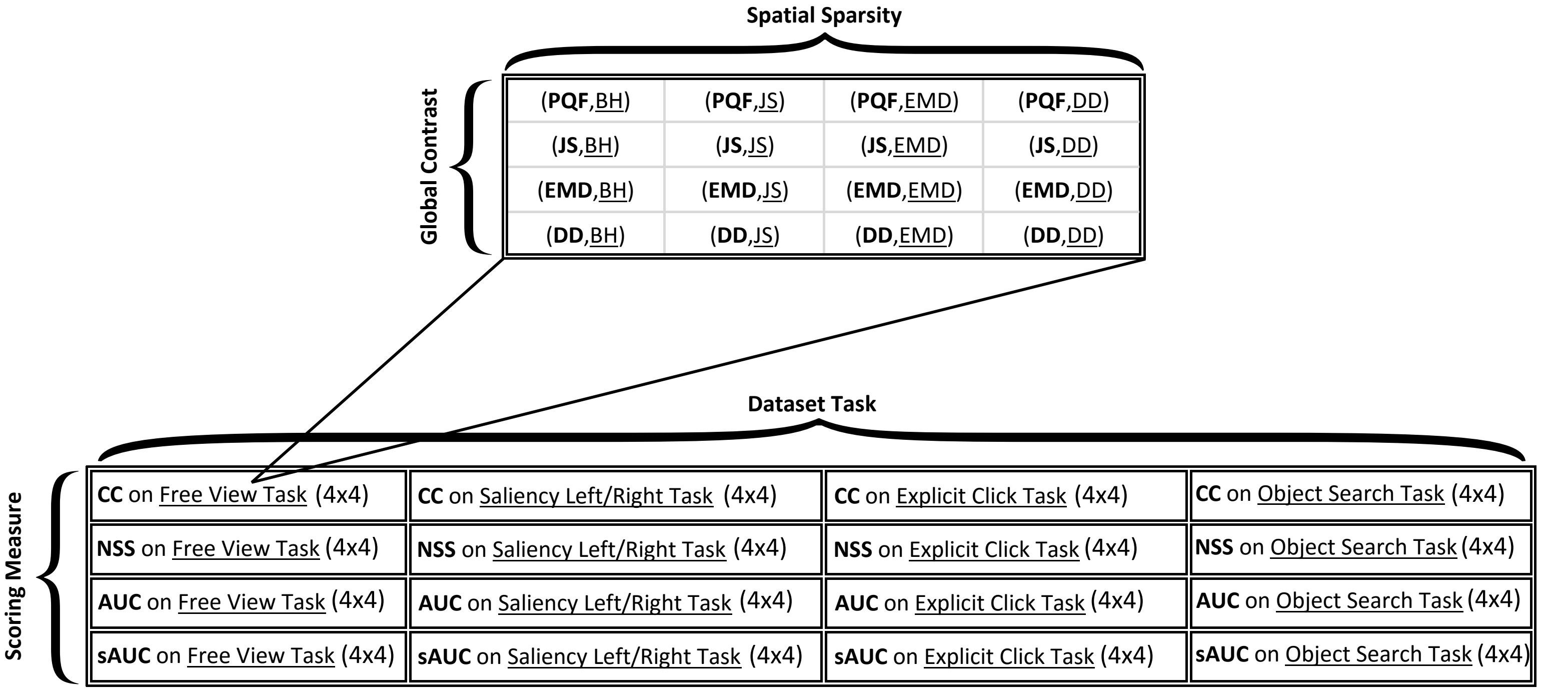
For example, assuming we wish to find the AUC scores computed on the explicit-click task for the equi-width histogram construction method. Using Table 4.6.1 as a reference, we see that the results lie in the third row and third column of the lower grid. This provides a 4×4 grid of results obtained by using all the combinations of global contrast and spatial sparsity distance measures, laid out as in the magnified grid.

4.6 Results

Following the presentation format described in Section §4.5, the results for this experiment are displayed in Table 4.6.2.

Looking at the results, the highest score in each grouping indicates that the original saliency formulation $S_S(\mathrm{PQF}, \mathrm{BH})$ (position $(1, 1)$ in each grouping) outperforms most of the other combinations, meaning that using different histogram distances is actually detrimental to the saliency scores. To further understand this, Figures 4.6.1 and 4.6.2 show two example images and the 16 different saliency maps generated using all the combinations of global contrast and spatial sparsity, indexed using Table 4.6.1. What immediately stands out is the contrast in the saliency maps. The first two columns, which correspond to bin-to-bin type spatial sparsity distances (BH and JS), have a much lower contrast saliency map than the last two columns, which are the cross-bin type distances (EMD and DD). This can be seen especially well in Figure 4.6.2 where there is a medium saliency value that permeates the images in the first two columns, with much lower background saliency values in the last two columns. This is a result of the spatial sparsity calculation. The cross-bin distances provide a much more continuous distance measure between the colour histograms, whereas the bin-to-bin distances require the bins to intersect. Histograms tend to become sparser as the dimensions increase due

Table 4.6.1: Grid used to index into the results tables. The 4×4 bordered cell shown magnified at the top represents the scores for a particular scoring measure and dataset task, and contains the scores using each of the global contrast distance measures (rows) and each of the spatial sparsity distance measures (columns). In the larger 4×4 grid at the bottom, each row represents a constant scoring measure and each column represents a constant dataset task. For clarity, bold items represent rows and underlined items represent columns.



Spatial Sparsity

Global Contrast

(**PQF**,BH)	(**PQF**,JS)	(**PQF**,EMD)	(**PQF**,DD)
(**JS**,BH)	(**JS**,JS)	(**JS**,EMD)	(**JS**,DD)
(**EMD**,BH)	(**EMD**,JS)	(**EMD**,EMD)	(**EMD**,DD)
(**DD**,BH)	(**DD**,JS)	(**DD**,EMD)	(**DD**,DD)

Dataset Task

Scoring Measure

CC on Free View Task (4x4)	**CC** on Saliency Left/Right Task (4x4)	**CC** on Explicit Click Task (4x4)	**CC** on Object Search Task (4x4)
NSS on Free View Task (4x4)	**NSS** on Saliency Left/Right Task (4x4)	**NSS** on Explicit Click Task (4x4)	**NSS** on Object Search Task (4x4)
AUC on Free View Task (4x4)	**AUC** on Saliency Left/Right Task (4x4)	**AUC** on Explicit Click Task (4x4)	**AUC** on Object Search Task (4x4)
sAUC on Free View Task (4x4)	**sAUC** on Saliency Left/Right Task (4x4)	**sAUC** on Explicit Click Task (4x4)	**sAUC** on Object Search Task (4x4)

Table 4.6.2: Equi-width histogram results (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.440	0.382	0.236	0.372	0.437	0.385	0.237	0.367	**0.350**	0.318	0.197	0.280	0.397	0.344	0.219	0.351
	0.449	0.387	0.221	0.365	0.441	0.384	0.217	0.355	0.331	0.295	0.164	0.253	0.421	0.363	0.217	0.356
	0.367	0.316	0.190	0.324	0.369	0.323	0.193	0.322	0.312	0.283	0.173	0.257	0.327	0.279	0.174	0.302
	0.457	0.391	0.217	0.364	**0.447**	0.387	0.213	0.353	0.335	0.294	0.159	0.248	**0.428**	0.367	0.213	0.356
NSS	0.714	0.624	0.383	0.598	**0.726**	0.646	0.407	0.604	**1.055**	0.996	0.717	0.799	0.659	0.572	0.370	0.583
	0.717	0.619	0.347	0.575	0.719	0.633	0.357	0.571	0.910	0.846	0.514	0.631	0.693	0.597	0.358	0.583
	0.588	0.508	0.301	0.511	0.610	0.540	0.331	0.527	0.970	0.918	0.654	0.760	0.526	0.450	0.282	0.489
	0.726	0.622	0.336	0.571	0.724	0.631	0.345	0.563	0.891	0.803	0.485	0.592	**0.697**	0.599	0.347	0.579
AUC	**0.692**	0.667	0.601	0.663	**0.691**	0.667	0.601	0.663	**0.719**	0.701	0.635	0.681	0.680	0.654	0.600	0.660
	0.686	0.657	0.588	0.655	0.686	0.659	0.588	0.654	0.701	0.679	0.609	0.661	0.679	0.651	0.593	0.656
	0.652	0.628	0.574	0.636	0.653	0.632	0.576	0.638	0.688	0.669	0.613	0.665	0.637	0.614	0.570	0.631
	0.691	0.659	0.586	0.656	0.688	0.659	0.586	0.653	0.697	0.670	0.604	0.653	**0.682**	0.653	0.589	0.657
sAUC	**0.542**	0.538	0.521	0.523	**0.549**	0.545	0.529	0.531	**0.614**	0.611	0.583	0.580	**0.530**	0.525	0.520	0.519
	0.524	0.520	0.510	0.511	0.532	0.529	0.516	0.517	0.583	0.580	0.556	0.551	0.518	0.514	0.511	0.509
	0.524	0.520	0.509	0.510	0.532	0.531	0.516	0.520	0.596	0.592	0.567	0.572	0.511	0.509	0.505	0.507
	0.523	0.517	0.505	0.508	0.528	0.524	0.511	0.512	0.574	0.566	0.548	0.539	0.514	0.510	0.504	0.505

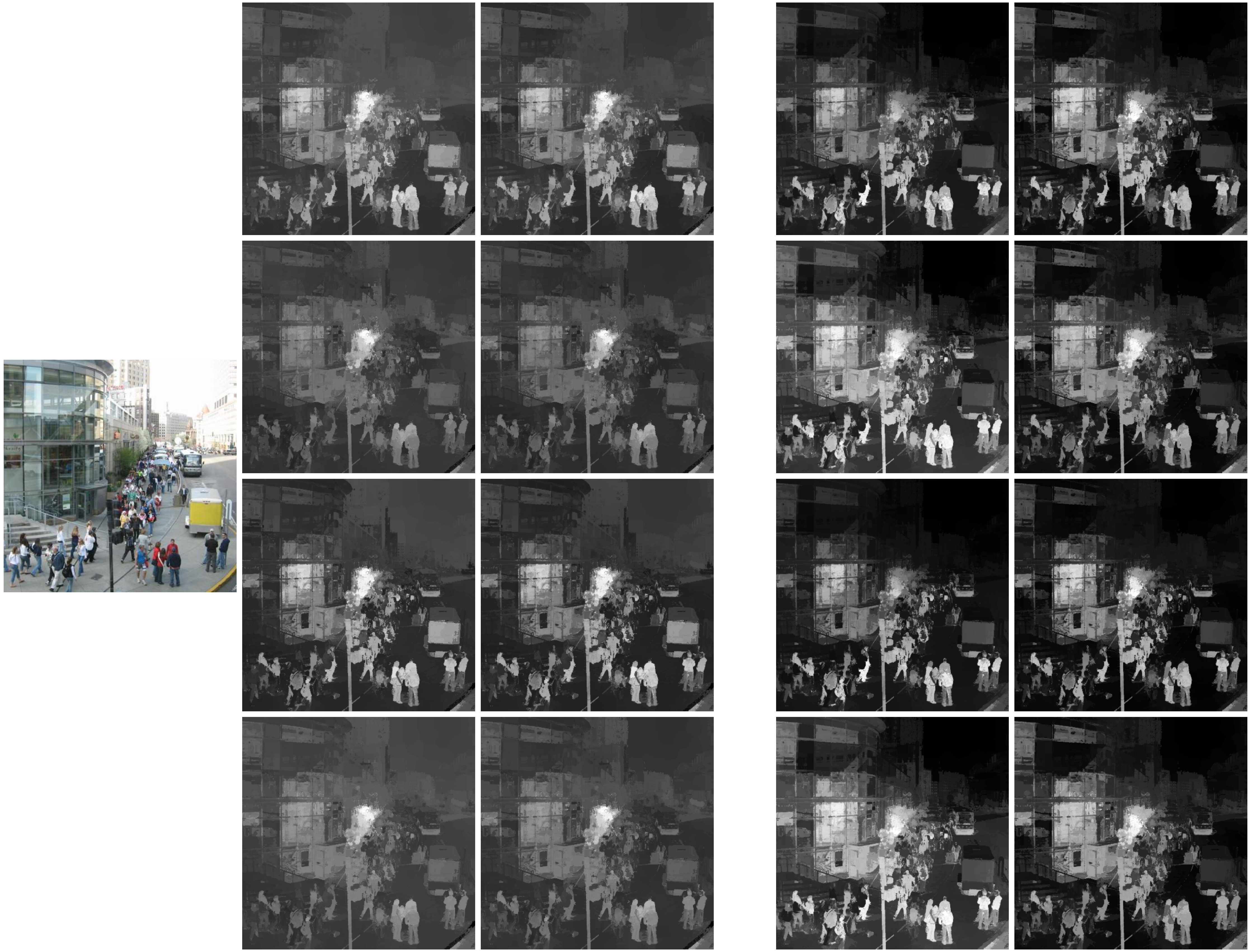


Figure 4.6.1: Example image with its 16 saliency maps generated using combinations of global contrast and spatial sparsity. Indexed as defined in Table 4.6.1.

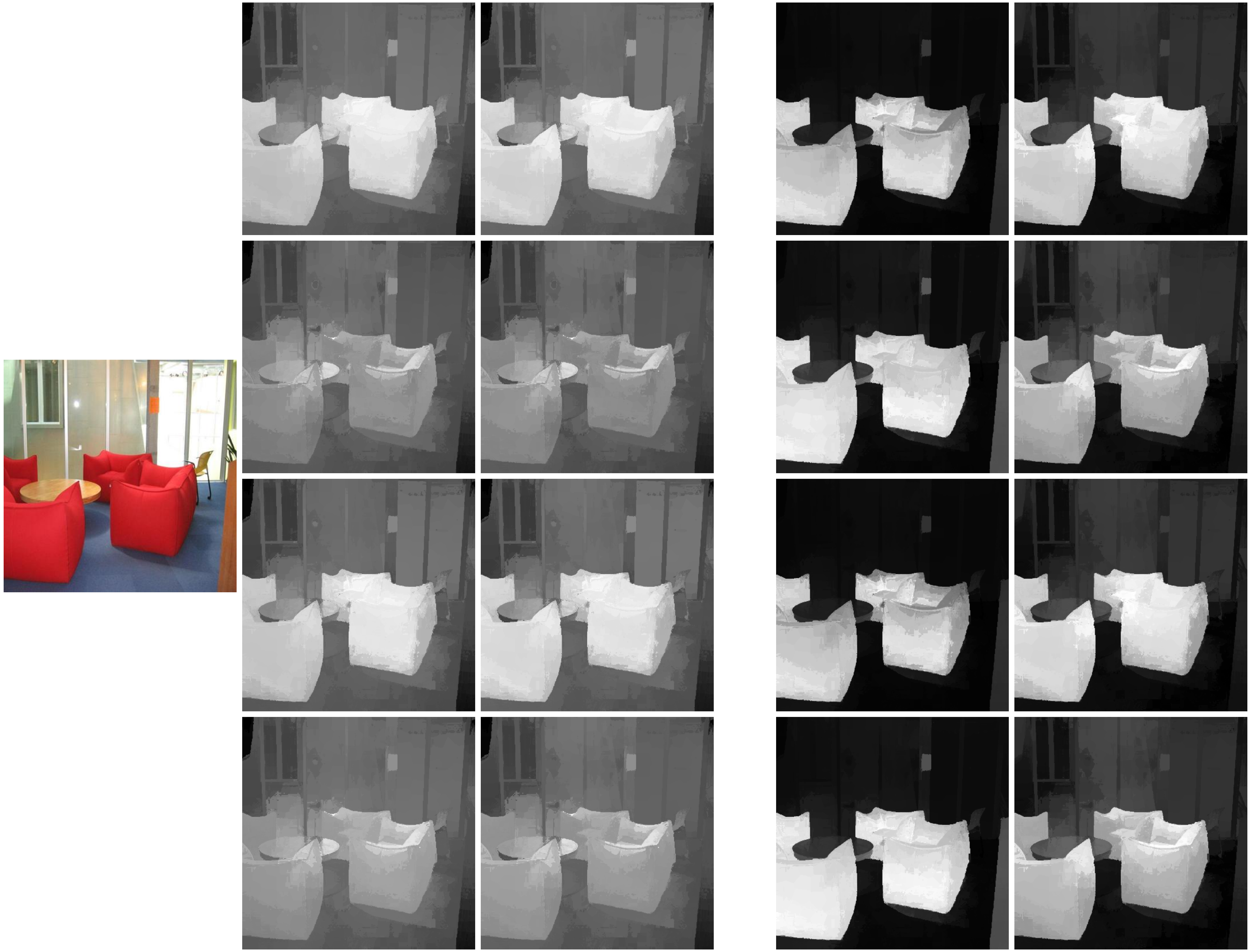


Figure 4.6.2: Another example image with its 16 saliency maps generated using combinations of global contrast and spatial sparsity. Indexed as defined in Table 4.6.1.

to the exponential increase in number of bins. This means the bin-to-bin distances become increasingly less adept at generating valid distances. The histograms could be shifted one bin over or an entire histogram width away, and still produce the same distance value. The spatial sparsity (Eq. 2.4.4) uses the distances to determine how prevalent and spread out the colours are in the image. If superpixels are being assigned similar similarity scores regardless of their actual distance in the colour space, they will generate largely the same saliency score, hence the permeating grey background.

Subjectively, it appears that the cross-bin distances improve the saliency maps, so to understand why the scoring would be lower we need to look at the fixations and click locations in Figure 4.6.3, which shows the same two example images overlaid with their fixation and click locations alongside the heatmaps generated from those locations. A human eye makes a saccade 3 to 4 times per second and it is unlikely that every saccade will land on a salient region throughout the image. This is especially due to the viewing strategy found in human observers coupled with us having our own ulterior motives and goals, even when free-viewing an image. This results in the eye-tracking data being quite haphazard and spread out around the image. For the bin-to-bin distances, these spread out fixations are still falling in relatively salient regions, which are contributing to the scores, albeit marginally, whereas with the cross-bin distances the background is being more effectively suppressed, which means the fixations falling there are actually detracting from the scores. Take the AUC measure for example, discussed in Section §2.5, which slides a discrimination threshold over the model generated saliency map, generating many thresholded saliency maps and computing the area under a precision versus recall graph. Due to the permeating saliency, the bin-to-bin distance maps will have more fixations in the thresholded area for longer, thereby increasing the AUC score.

Interestingly, most combinations score relatively higher for the explicit-click task, except for the CC scoring measure. Going back to Figure 4.6.3, the third row represents the explicit-click task. It is remarkable how tight and consistent the participants' choices are, here as well as throughout the dataset. Because they keep picking the same object, which usually involves a salient colour or shape, the saliency maps generally predict that selection quite well. This tightness of click locations is also why the CC score drops; due to the click locations heatmap becoming very peaky, it does not correlate very well with the generated saliency maps. By giving the participants an explicit goal, as well as the autonomy and freedom to choose when their choices are made, their top-down and bottom-up goals become aligned to detect saliency, thereby providing a better assessment of the saliency in the image. It is posited that allowing participants to choose up to n salient regions in the image within a certain time frame would produce similar saliency maps to those created by the cross-bin histogram distances. That is unfortunately out of the scope of this dissertation and will be left for future work.

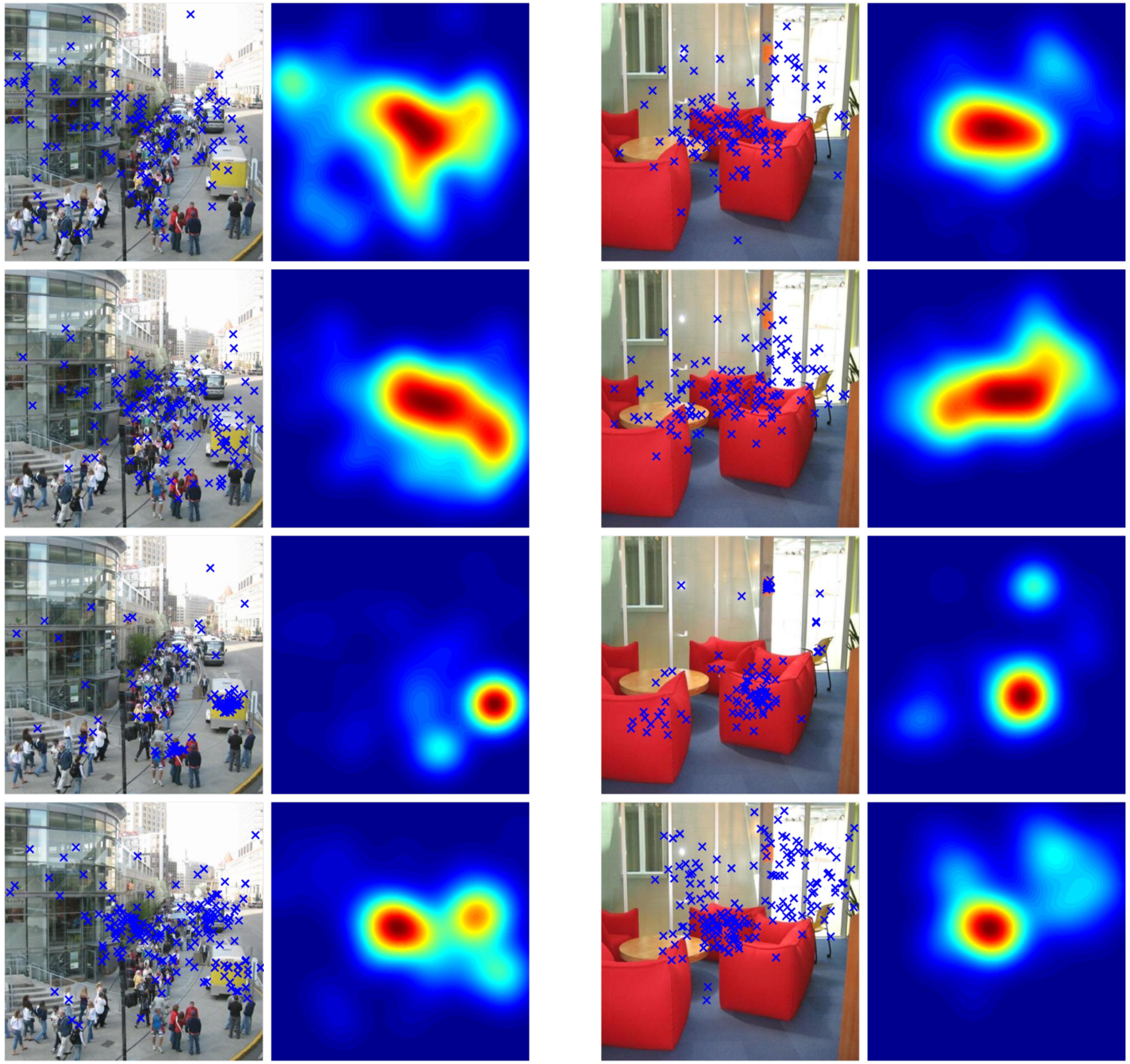


Figure 4.6.3: Two example images overlaid with their eye-tracking and click data for all 4 dataset tasks. From top to bottom: free-view, saliency left/right, explicit-click, object search.

4.7 Conclusion

The goal of this experiment was to establish whether different histogram distance measures used in a colour saliency algorithm would affect its performance. Saliency maps were generated using combinations of histogram distances and measured on a database of four viewing tasks. Out of all the histogram distance combinations, the original saliency equations scored highest on the most metrics, but, perceptually, the cross-bin distances seem to generate better saliency maps. A reason for the scoring disparity was provided on the basis of the bin-to-bin distances not being able to adequately suppress background regions, which in turn map better to the more spread out eye-tracking fixation locations. It was observed that the cross-bin type methods generated higher-contrast saliency maps, and it is this higher accuracy that lowered the scores due to the random and haphazard eye movements tracked during tasks. Based on the explicit-click task and how well it matches perception, a new experiment is proposed which could improve the current saliency scoring mechanisms: participants are asked to select up to some small n of the most salient regions in an image, perhaps under a time limit, instead of having their eyes tracked. The act of using a mouse is more directed and precise than eye-tracking data, which also aligns the goal-driven top-down and automatic, feature-driven bottom-up goals to isolate saliency detection from any intrinsic free-viewing goals.

Chapter 5
Experiment 3: The Effects of Histogram Construction Methods and Distances on Saliency Detection

The previous two experiments look at the impact the histogram construction method and the histogram distances have on saliency detection in isolation. The purpose of this experiment is to run through all combinations of histogram construction methods and histogram distances to determine what combined effect they might have. The experiment is laid out as follows: the aim is clarified in Section §5.1, the hypothesis is stated in Section §5.2 and the apparatus used is detailed in Section §5.3. The experimental method is provided in Section §5.4, a reference to how the results are presented is given in Section §5.5, the results are presented in Section §5.6 and the conclusions drawn are discussed in Section §5.7.

5.1 Aim

The aim of this experiment is to determine the effects of combining various histogram construction methods with histogram distances on saliency detection.

5.2 Hypothesis

Due to some improvements from the prior experiments in isolation, it is hypothesized that a combination of histogram construction methods and histogram distances used in the saliency calculation will improve the results even further.

5.3 Apparatus

The experiment was conducted on an Asus UX303LN laptop computer, containing an Intel i7-4510U 2.0GHz CPU, 12GB RAM and 256GB SSD HDD, running MATLAB

R2013a on Microsoft Windows 8.1.

5.4 Method

This experiment will use the methods from both previous experiments, Section §3.4 and Section §4.4. For this experiment, 4 metrics (CC, NSS, AUC, sAUC) from Section §2.5 are computed across the 4 database tasks detailed in Section §2.5.1, using the 3 histogram construction methods introduced in Section §2.2, with 4 global contrast distances and 4 spatial sparsity distances explained in Section §2.3.

5.5 Presentation of Results

The results for this experiment are presented in the same fashion as in Section §4.5. Table 4.6.1 is again used to index into the results tables, the only difference being that there will now be 3 tables of 256 scores, one for each histogram construction method.

5.6 Results

The results for all the equi-width combinations are the same as in Section §4.6 and are repeated in Table 5.6.1 for convenience, the equi-depth combination results are found in Table 5.6.2 and the diagonal histogram combination results are found in Table 5.6.3. As discovered and discussed in Experiment 1, there is a slight improvement moving from equi-width to equi-depth histogram construction methods for most tasks, with the d-hist method more-or-less falling in-between them. Interestingly, it seems to be the opposite for the explicit-click task. To illustrate this point, the percentage increase (or decrease) in moving from the equi-width to the equi-depth histograms is shown in Table 5.6.4, with green indicating that using equi-depth histograms improved scores and red indicating that they worsened scores. A possible cause of the worsening scores for the explicit-click task is that the equi-depth histogram tends to flatten out the low-density regions, pooling them together and thereby reducing their overall peak saliency, which when coupled with the highly accurate and selective explicit-click task lowers the resulting scores. This might also explain the reason for the increase in scores for the eye-tracking tasks: due to the fixation locations being more spread out, the scores benefit from having a more distributed saliency map.

Table 5.6.1: Equi-width histogram results (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.440	0.382	0.236	0.372	0.437	0.385	0.237	0.367	**0.350**	0.318	0.197	0.280	0.397	0.344	0.219	0.351
	0.449	0.387	0.221	0.365	0.441	0.384	0.217	0.355	0.331	0.295	0.164	0.253	0.421	0.363	0.217	0.356
	0.367	0.316	0.190	0.324	0.369	0.323	0.193	0.322	0.312	0.283	0.173	0.257	0.327	0.279	0.174	0.302
	0.457	0.391	0.217	0.364	**0.447**	0.387	0.213	0.353	0.335	0.294	0.159	0.248	**0.428**	0.367	0.213	0.356
NSS	0.714	0.624	0.383	0.598	**0.726**	0.646	0.407	0.604	**1.055**	0.996	0.717	0.799	0.659	0.572	0.370	0.583
	0.717	0.619	0.347	0.575	0.719	0.633	0.357	0.571	0.910	0.846	0.514	0.631	0.693	0.597	0.358	0.583
	0.588	0.508	0.301	0.511	0.610	0.540	0.331	0.527	0.970	0.918	0.654	0.760	0.526	0.450	0.282	0.489
	0.726	0.622	0.336	0.571	0.724	0.631	0.345	0.563	0.891	0.803	0.485	0.592	**0.697**	0.599	0.347	0.579
AUC	**0.692**	0.667	0.601	0.663	**0.691**	0.667	0.601	0.663	**0.719**	0.701	0.635	0.681	0.680	0.654	0.600	0.660
	0.686	0.657	0.588	0.655	0.686	0.659	0.588	0.654	0.701	0.679	0.609	0.661	0.679	0.651	0.593	0.656
	0.652	0.628	0.574	0.636	0.653	0.632	0.576	0.638	0.688	0.669	0.613	0.665	0.637	0.614	0.570	0.631
	0.691	0.659	0.586	0.656	0.688	0.659	0.586	0.653	0.697	0.670	0.604	0.653	**0.682**	0.653	0.589	0.657
sAUC	**0.542**	0.538	0.521	0.523	**0.549**	0.545	0.529	0.531	**0.614**	0.611	0.583	0.580	**0.530**	0.525	0.520	0.519
	0.524	0.520	0.510	0.511	0.532	0.529	0.516	0.517	0.583	0.580	0.556	0.551	0.518	0.514	0.511	0.509
	0.524	0.520	0.509	0.510	0.532	0.531	0.516	0.520	0.596	0.592	0.567	0.572	0.511	0.509	0.505	0.507
	0.523	0.517	0.505	0.508	0.528	0.524	0.511	0.512	0.574	0.566	0.548	0.539	0.514	0.510	0.504	0.505

Table 5.6.2: Equi-depth histogram results (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.490	0.446	0.236	0.435	0.481	0.442	0.235	0.422	**0.363**	0.338	0.187	0.300	0.454	0.414	0.222	0.421
	0.516	0.470	0.223	0.434	0.490	0.448	0.211	0.406	0.321	0.294	0.133	0.249	**0.510**	0.467	0.230	0.441
	0.442	0.400	0.207	0.402	0.434	0.397	0.207	0.391	0.331	0.306	0.169	0.280	0.407	0.368	0.192	0.386
	0.516	0.454	0.217	0.420	**0.493**	0.430	0.209	0.391	0.335	0.275	0.141	0.234	0.505	0.453	0.222	0.429
NSS	0.783	0.713	0.375	0.687	0.787	0.728	0.396	0.675	**0.949**	0.897	0.608	0.695	0.751	0.683	0.371	0.692
	0.810	0.737	0.341	0.672	0.781	0.718	0.335	0.631	0.706	0.652	0.330	0.455	**0.836**	0.765	0.376	0.720
	0.698	0.630	0.320	0.628	0.707	0.650	0.346	0.622	0.870	0.820	0.560	0.653	0.657	0.593	0.309	0.623
	0.810	0.705	0.333	0.645	**0.789**	0.679	0.334	0.600	0.776	0.568	0.381	0.393	0.825	0.734	0.360	0.693
AUC	0.715	0.694	0.600	0.692	0.713	0.695	0.602	0.688	**0.725**	0.709	0.629	0.686	0.706	0.687	0.602	0.692
	0.720	0.698	0.593	0.688	0.713	0.693	0.589	0.678	0.688	0.671	0.583	0.637	**0.719**	0.699	0.603	0.696
	0.688	0.669	0.581	0.673	0.688	0.670	0.582	0.671	0.697	0.680	0.606	0.670	0.676	0.659	0.579	0.671
	0.719	0.691	0.587	0.681	**0.713**	0.684	0.584	0.670	0.698	0.651	0.589	0.618	0.716	0.692	0.597	0.688
sAUC	**0.531**	0.528	0.518	0.513	**0.541**	0.539	0.526	0.520	**0.592**	0.588	0.574	0.549	**0.524**	0.521	0.518	0.514
	0.501	0.499	0.498	0.485	0.507	0.505	0.501	0.485	0.520	0.520	0.517	0.467	0.504	0.501	0.505	0.495
	0.523	0.519	0.505	0.507	0.531	0.529	0.513	0.513	0.574	0.569	0.555	0.541	0.512	0.511	0.505	0.506
	0.503	0.489	0.496	0.478	0.509	0.493	0.500	0.476	0.534	0.494	0.528	0.446	0.502	0.493	0.503	0.488

Table 5.6.3: Diagonal histogram results (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.472	0.423	0.235	0.410	0.466	0.422	0.235	0.402	**0.362**	0.335	0.194	0.296	0.433	0.388	0.218	0.392
	0.491	0.440	0.219	0.404	0.471	0.425	0.210	0.384	0.326	0.297	0.142	0.251	**0.478**	0.430	0.223	0.405
	0.401	0.357	0.197	0.365	0.399	0.360	0.199	0.361	0.322	0.298	0.177	0.276	0.362	0.322	0.179	0.345
	0.492	0.438	0.213	0.400	**0.472**	0.422	0.205	0.380	0.329	0.294	0.141	0.247	0.477	0.426	0.216	0.400
NSS	0.761	0.682	0.379	0.653	**0.766**	0.701	0.401	0.649	**1.051**	1.005	0.703	0.781	0.716	0.642	0.367	0.646
	0.776	0.694	0.336	0.630	0.758	0.688	0.337	0.604	0.810	0.761	0.408	0.536	**0.784**	0.704	0.364	0.661
	0.639	0.568	0.310	0.575	0.655	0.597	0.340	0.582	0.966	0.926	0.667	0.760	0.587	0.521	0.292	0.560
	0.774	0.686	0.325	0.619	0.756	0.678	0.325	0.593	0.805	0.723	0.398	0.505	0.776	0.692	0.348	0.649
AUC	0.707	0.683	0.599	0.681	**0.705**	0.684	0.601	0.677	**0.727**	0.711	0.634	0.689	0.696	0.674	0.599	0.679
	0.708	0.683	0.589	0.675	0.703	0.681	0.587	0.669	0.697	0.680	0.591	0.650	**0.705**	0.683	0.597	0.680
	0.669	0.649	0.575	0.657	0.669	0.652	0.577	0.656	0.696	0.682	0.614	0.674	0.657	0.638	0.573	0.653
	0.709	0.683	0.584	0.672	0.704	0.679	0.582	0.665	0.695	0.671	0.587	0.641	0.704	0.681	0.592	0.678
sAUC	**0.534**	0.530	0.519	0.517	**0.542**	0.540	0.528	0.524	**0.606**	0.604	0.582	0.568	**0.525**	0.522	0.518	0.516
	0.507	0.505	0.500	0.496	0.513	0.512	0.505	0.498	0.548	0.548	0.532	0.510	0.506	0.505	0.503	0.500
	0.524	0.520	0.511	0.511	0.531	0.530	0.517	0.519	0.589	0.589	0.568	0.565	0.514	0.511	0.509	0.509
	0.506	0.501	0.495	0.491	0.512	0.506	0.501	0.494	0.545	0.536	0.527	0.498	0.504	0.500	0.499	0.496

Table 5.6.4: Percentage increase (or decrease) in scores when using equi-depth histograms over equi-width histograms. Green indicates that using equi-depth histograms improved scores and red indicates that they worsened scores.

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	11%	17%	0%	17%	10%	15%	-1%	15%	4%	6%	-5%	7%	14%	20%	1%	20%
	15%	22%	1%	19%	11%	17%	-3%	14%	-3%	0%	-19%	-2%	21%	29%	6%	24%
	20%	26%	9%	24%	18%	23%	7%	21%	6%	8%	-2%	9%	25%	32%	11%	28%
	13%	16%	0%	15%	10%	11%	-2%	11%	0%	-6%	-11%	-6%	18%	23%	4%	21%
NSS	10%	14%	-2%	15%	8%	13%	-3%	12%	-10%	-10%	-15%	-13%	14%	19%	0%	19%
	13%	19%	-2%	17%	9%	13%	-6%	11%	-22%	-23%	-36%	-28%	21%	28%	5%	24%
	19%	24%	6%	23%	16%	20%	5%	18%	-10%	-11%	-14%	-14%	25%	32%	10%	28%
	12%	13%	-1%	13%	9%	8%	-3%	7%	-13%	-29%	-21%	-34%	18%	23%	4%	20%
AUC	3%	4%	0%	4%	3%	4%	0%	4%	1%	1%	-1%	1%	4%	5%	0%	5%
	5%	6%	1%	5%	4%	5%	0%	4%	-2%	-1%	-4%	-4%	6%	7%	2%	6%
	5%	7%	1%	6%	5%	6%	1%	5%	1%	2%	-1%	1%	6%	7%	2%	6%
	4%	5%	0%	4%	4%	4%	0%	2%	0%	-3%	-2%	-5%	5%	6%	1%	5%
sAUC	-2%	-2%	-1%	-2%	-2%	-1%	-1%	-2%	-4%	-4%	-2%	-5%	-1%	-1%	0%	-1%
	-4%	-4%	-2%	-5%	-5%	-5%	-3%	-6%	-11%	-10%	-7%	-15%	-3%	-3%	-1%	-3%
	0%	0%	-1%	-1%	0%	0%	-1%	-1%	-4%	-4%	-2%	-5%	0%	0%	0%	0%
	-4%	-5%	-2%	-6%	-4%	-6%	-2%	-7%	-7%	-13%	-4%	-17%	-2%	-3%	0%	-4%

5.7 Conclusion

This experiment was conducted to study the effects of all the combinations of histogram construction methods and histogram distances on saliency detection. It was confirmed that using equi-depth histograms can improve saliency scores, but only for eye-tracking tasks, and actually decreases the scores for the explicit-click saliency task. This is mainly attributed to the equi-depth histogram's spreading out of the saliency map due to it lumping low-density regions together, thereby reducing the overall saliency. This fact aids the eye-tracking tasks because eye-tracking data is much more random and distributed over the image, whereas the explicit-click task is more precise and focused. This again points to a problem with the scoring mechanisms, as the saliency maps generated by combinations with lower scores are more aligned with what seems perceptually salient. Additional selections for the explicit-click task will allow a more complete view of the salient regions in an image, and will not unnecessarily penalise the higher contrast saliency maps.

It was observed that the eye-tracking data and even some of the click data is quite random and spread out around the image in a somewhat haphazard manner, whereas the saliency maps generated have pixel-level accuracy and retain object boundaries relatively well. It was questioned as to what would happen to the saliency scores if the saliency maps were smoothed slightly so as to better accommodate the randomness in the data. This auxiliary experiment is not critical to the dissertation and can be found in Appendix A, but an interesting finding is how much the scores increase simply by smoothing the saliency maps with a Gaussian filter with $\sigma = 0.5°$ of visual angle, as can be seen in Table A.5. This reinforces the idea that the scoring mechanisms are highly susceptible to variations and new and more robust methods are needed.

Chapter 6
Experiment 4: Removing the Center Bias

During the process of working with the cross-bin histograms and noticing that the spatial sparsity (Eq. 2.4.4) has a built-in center bias, it was investigated whether there could be another way to compute the spatial sparsity of a superpixel without introducing a heavy center bias. The aim, apparatus and method of the experiment are stated in Sections §6.1, §6.2 and §6.3 respectively, while the results and conclusions are presented in Sections §6.4 and §6.5 respectively.

6.1 Aim

The aim of this experiment is to determine whether a novel computation for the spatial sparsity measure can be developed without a center bias.

6.2 Apparatus

The experiment was conducted on an Asus UX303LN laptop computer, containing an Intel i7-4510U 2.0GHz CPU, 12GB RAM and 256GB SSD HDD, running MATLAB R2013a on Microsoft Windows 8.1.

6.3 Method

The assumption that the spatial sparsity measure is working on is that salient objects' colours are sparsely distributed around the frame with the converse being that background and non-salient objects' colour distributions are more densely distributed around the frame. Liu et al. [33] realise this assumption by calculating a sum of distances from all superpixel centroids to the center of the frame, weighted by how similar and nearby the superpixels are to the superpixel in question (Eq. 2.4.3). This, in effect, does what is intended by giving an indication of the spread of a certain colour distribution in the image, but has the negative effect of introducing a center bias. This center bias improves saliency scores but generates less perceptually-meaningful saliency maps.

While experimenting with the EMD between superpixel colour distributions, it was noted that the EMD distances between each superpixel are much more continuous than their bin-to-bin counterpart, the Bhattacharyya coefficient. Thus, it was easier to gain access to very similar colour distributions in the frame, based on perceptual distance rather than bin overlap. A method was developed which, for every superpixel, would compute the EMD between its and every other superpixel's colour distribution. A similarity measure is then obtained by inverse normalising these distances. Superpixels with similar colour distributions are identified by thresholding the similarity values. The spatial distribution of each superpixel is then computed as the sum of distances from the joint centroid of all superpixels with a similar colour distribution to the one in question:

$$\mathrm{D_{EMD}}(sp_i, sp_j) = \mathrm{EMD}(\mathrm{CH}_i, \mathrm{CH}_j), \tag{6.3.1}$$

$$\lambda_{\mathrm{intra_{EMD}}}(sp_i, sp_j) = \frac{\max\left[\mathrm{D_{EMD}}(sp)\right] - \mathrm{D_{EMD}}(sp_i, sp_j)}{\max\left[\mathrm{D_{EMD}}(sp)\right] - \min\left[\mathrm{D_{EMD}}(sp)\right]}, \tag{6.3.2}$$

$$\mathrm{SD}(sp_i) = \sum_j \left\|\mu_j - \mu_c\right\|_2, \tag{6.3.3}$$

where μ_j is the centroid of a superpixel with $\lambda_{\mathrm{intra_{EMD}}}(sp_i, sp_j) \geq t$ and μ_c is the joint centroid of all the superpixels with similarity values above the threshold t. The final spatial sparsity is calculated as an inverse normalisation, as in Eq. 2.4.4. This is equivalent to switching the center of the frame in the original formula to the mean location of a superpixel's most similar superpixels. If the superpixel has similar superpixels spread around the frame, then μ_c approaches the center of the frame and it works much like the original equation. However, if the superpixel has similar superpixels that are tightly bunched, μ_c will be close to all of them and results in a small spatial distribution sum as expected. In the experiments $t = 0.9$ produced satisfactory results, and corresponds to patches being "90% similar in terms of colour" as defined by the EMD distance.

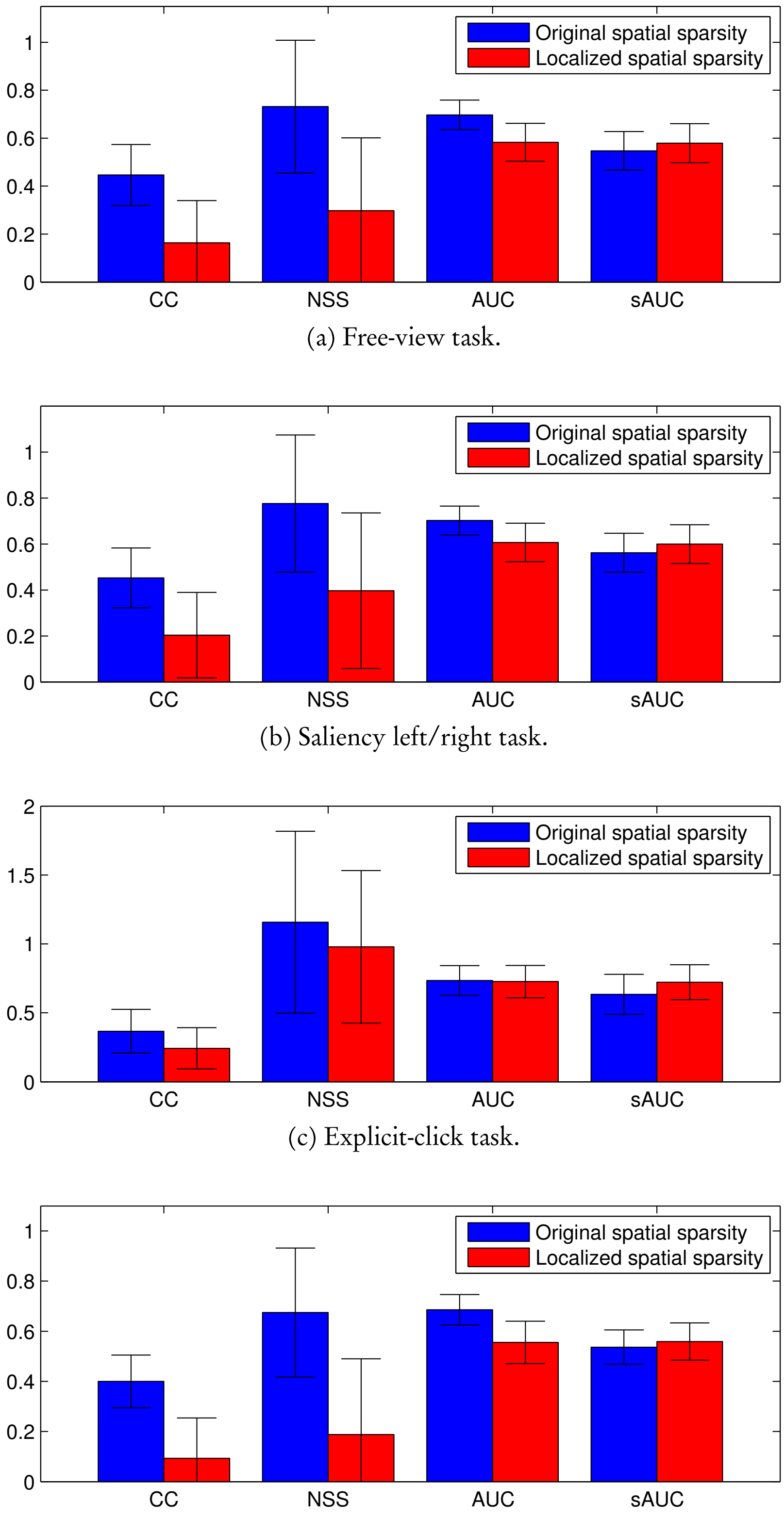
6.4 Results

Removing the center bias drastically lowers most of the scores, as can be seen in the first 3 scores of Figure 6.4.1. What is surprising is the increase in sAUC across all 4 dataset tasks. The sAUC measure was designed to cater for the center bias, which would naturally lower the original method's AUC score, as is clearly seen. The fact that the newly-proposed method is unaffected by using sAUC as opposed to AUC indicates that it has successfully removed the center bias from the saliency calculation.

Figure 6.5.1 shows some example images with the saliency maps generated using the original method and the proposed method. The new method reduces the center bias and better implements the assumption of spatial sparsity provided in Liu et al. [33]. The third row makes this very clear by firstly removing the center bias, which can be seen in the original map as a permeating saliency value increasing towards the center. Secondly, it enhances the no entry sign's saliency value (because it is the only red object in the scene) and lowers the lamp posts' saliency values, because although they are fairly salient as individuals they are spread out across the image and are therefore not quite as salient.

6.5 Conclusion

This experiment was conducted to reduce the center bias present in the colour saliency calculation found in Liu et al. [33]. The proposed method removes the center bias by moving the frame-center fixed-point reference from the original equation to the mean position of very similar superpixels, which are calculated using the more continuous cross-bin histogram distances. It is shown that the method does indeed remove the center bias by showing invariance in the AUC and sAUC scores generated from the dataset as well as from visual inspection. The removal of the center bias significantly affects the CC, NSS and AUC scores, reiterating how susceptible these scoring methods are to the center bias.



(a) Free-view task.

(b) Saliency left/right task.

(c) Explicit-click task.

(d) Object proposal task.

Figure 6.4.1: Results of using the newly proposed spatial sparsity measure (error bars indicate one standard deviation).

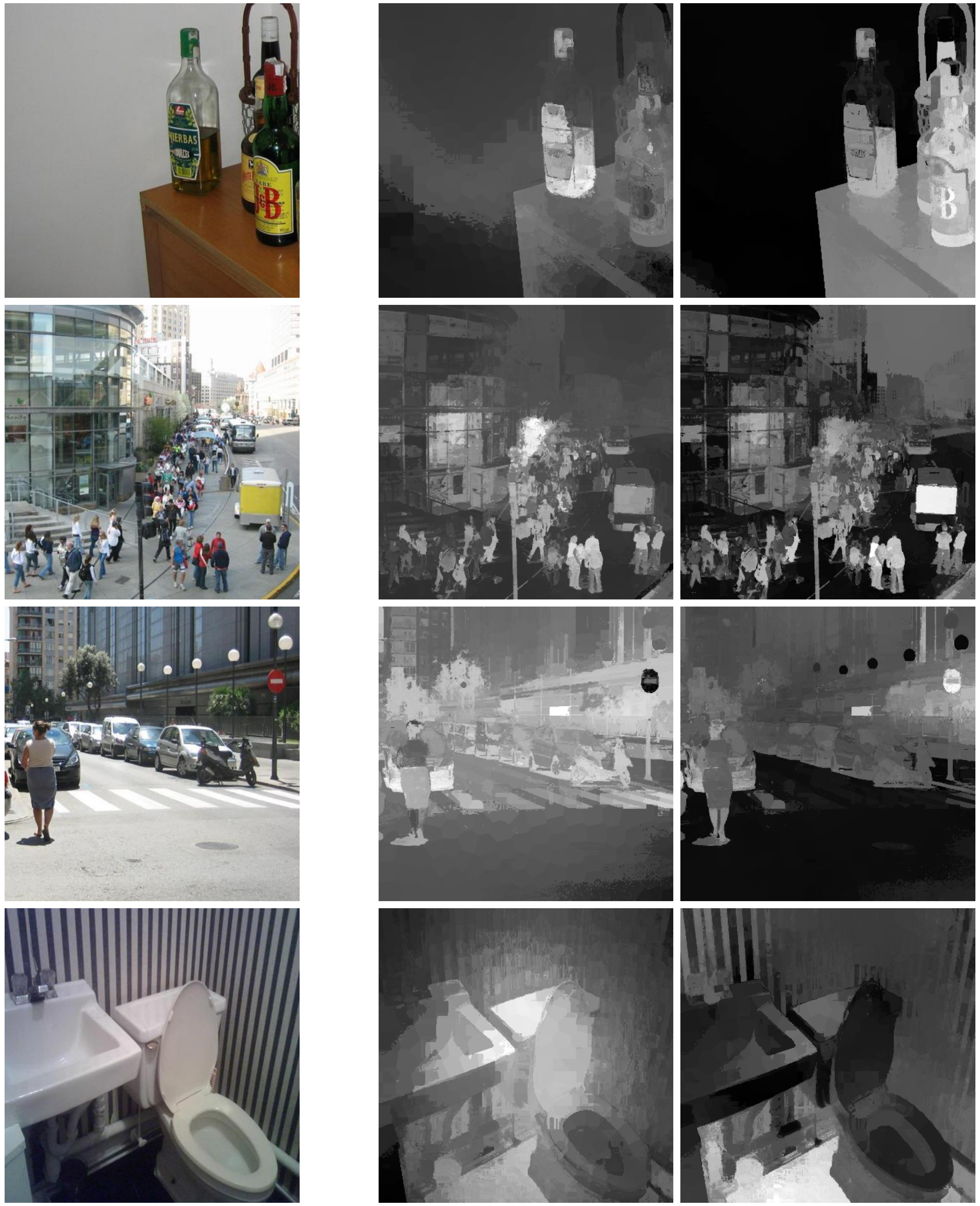


Figure 6.5.1: Example images with their original saliency maps and saliency maps generated with the proposed method for spatial sparsity.

Chapter 7
Final Conclusions

Saliency research has seen a lot of activity in recent years, starting from a biological standpoint and migrating into the computational realms. Many models ranging from biologically plausible, to information theoretic, to statistical in nature have been proposed, each contributing to the overarching field of visual saliency. This dissertation focuses on a particular statistical saliency model which computes the saliency at a region by comparing the local region's colour distribution with the global frame's colour distribution, and assuming that salient objects are local and sparsely distributed across the frame. The original model achieves state-of-the-art performance on existing datasets across most current metrics, which is why it was chosen as the focus.

The dissertation designed and implemented experiments to assess whether the histogram construction methods and histogram distances, which make up the crux of the algorithm, would affect the saliency scores it generated. It was shown that equi-depth histograms are better able to characterise the global frame colour distribution, which allows for better discovery of salient regions. It was also shown that cross-bin histogram distances create a perceptually more meaningful saliency map, but fall prey to the center and search bias present in eye-tracking saliency-scoring. The explicit saliency selection task is more indicative of the salient regions in the image, and future experiments requesting participants to select multiple salient regions is proposed. This would alleviate the random eye movements, as well as align the top-down and bottom-up tasks to provide a more meaningful saliency measure.

Future work might aim to improve the cross-bin distance computation times via parallel programming, or to take these findings into the spatiotemporal and depth-of-field realms. Vision is an extremely powerful and information-rich sensory input, which saliency allows us to make computationally tractable and deepens our knowledge of how we see and perceive the world.

Bibliography

[1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels," Tech. Rep., 2010.

[2] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, pp. 33–41, 1984.

[3] F. J. Aherne, N. A. Thacker, and P. I. Rockett, "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika*, vol. 34, no. 4, pp. 363–368, 1998.

[4] L. Bao, Q. Yang, and H. Jin, "Fast edge-preserving patchmatch for large displacement optical flow," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 4996–5006, 2014.

[5] C. Baumann, "Ewald Hering's opponent colors. History of an idea," *Der Ophthalmologe: Zeitschrift der Deutschen Ophthalmologischen Gesellschaft*, vol. 89, no. 3, pp. 249–252, 1992.

[6] R. Bellman, *Adaptive Control Processes: A Guided Tour.* Princeton University Press, 1961.

[7] A. Borji, H. R. Tavakoli, D. N. Sihite, and L. Itti, "Analysis of scores, datasets, and models in visual saliency prediction," in *IEEE International Conference on Computer Vision*, 2013, pp. 921–928.

[8] T. Brox, C. Bregler, and J. Malik, "Large displacement optical flow," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 41–48.

[9] N. D. Bruce and J. K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, p. 5, 2009.

[10] M. E. Celebi, "Improving the performance of k-means for color quantization," *Image and Vision Computing*, vol. 29, no. 4, pp. 260–271, 2011.

[11] S.-H. Cha, "Comprehensive survey on distance/similarity measures between probability density functions," in *International Journal of Mathematical Models and Methods in Applied Sciences.* Citeseer, 2007.

[12] C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human photoreceptor topography," *Journal of Comparative Neurology*, vol. 292, no. 4, pp. 497–523, 1990.

[13] L. Denby and C. Mallows, "Variations on the histogram," *Journal of Computational and Graphical Statistics*, vol. 18, no. 1, pp. 21–31, 2009.

[14] M. Frackiewicz and H. Palus, "Clustering with k-harmonic means applied to colour image quantization," in *IEEE International Symposium on Signal Processing and Information Technology*, 2008, pp. 52–57.

[15] D. Gao and N. Vasconcelos, "Decision-theoretic saliency: computational principles, biological plausibility, and implications for neurophysiology and psychophysics," *Neural Computation*, vol. 21, no. 1, pp. 239–271, 2009.

[16] J. Guild, "The colorimetric properties of the spectrum," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, pp. 149–187, 1932.

[17] J. E. Hoffman and B. Subramaniam, "The role of visual attention in saccadic eye movements," *Perception & Psychophysics*, vol. 57, no. 6, pp. 787–795, 1995.

[18] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[19] Y. Ioannidis, "The history of histograms (abridged)," in *Proceedings of the 29th International Conference on Very Large Databases*, vol. 29. VLDB Endowment, 2003, pp. 19–30.

[20] ISO 11664-1:2007, *Colorimetry – Part 1: CIE standard colorimetric observers*. ISO, Geneva, Switzerland.

[21] ISO 11664-4:2008, *Colorimetry – Part 4: CIE 1976 L*a*b* Colour space*. ISO, Geneva, Switzerland.

[22] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[23] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[24] W. James, *The Principles of Psychology*. Henry Holt & Co., 1890, vol. 1, no. 2, ch. XI, pp. 403–404.

[25] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *IEEE 12th International Conference on Computer Vision*, 2009, pp. 2106–2113.

[26] G. S. Kell, "Density, thermal expansivity, and compressibility of liquid water from 0° to 150°c: Correlations and tables for atmospheric pressure and saturation reviewed and expressed on 1968 temperature scale," *Journal of Chemical and Engineering Data*, vol. 20, no. 1, 1975.

[27] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry." *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.

[28] K. Koehler, F. Guo, S. Zhang, and M. P. Eckstein, "What do saliency models predict?" *Journal of Vision*, vol. 14, no. 3, p. 14, 2014.

[29] R. Kurzweil, *How to create a mind: The secret of human thought revealed.* Penguin, 2012, p. 95.

[30] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Methods*, vol. 45, no. 1, pp. 251–266, 2013.

[31] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 4, pp. 996–1010, 2013.

[32] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2006, pp. 246–253.

[33] Z. Liu, X. Zhang, S. Luo, and O. Le Meur, "Superpixel-based spatiotemporal saliency detection," *IEEE Transactions on Circuits and Systems for Video Technology*, p. 1, 2014.

[34] E. Niebur and C. Koch, "Control of selective visual attention: Modeling the "where" pathway," *Advances in Neural Information Processing Systems*, pp. 802–808, 1996.

[35] R. J. Peters, A. Iyer, L. Itti, and C. Koch, "Components of bottom-up gaze allocation in natural images," *Vision Research*, vol. 45, no. 18, pp. 2397–2416, 2005.

[36] G. Piatetsky-Shapiro and C. Connell, "Accurate estimation of the number of tuples satisfying a condition," in *ACM SIGMOD Record*, vol. 14, no. 2, 1984, pp. 256–276.

[37] D. Purves, D. Fitzpatrick, L. Katz, A. Lamantia, J. McNamara, S. Williams, and G. Augustine, *Neuroscience*. Sinauer Associates, 2001.

[38] Y. Rubner, "Perceptual metrics for image database navigation," Ph.D. dissertation, Stanford University, 1999.

[39] D. L. Ruderman, "The statistics of natural images," *Network: Computation in Neural Systems*, vol. 5, no. 4, pp. 517–548, 1994.

[40] D. D. Salvucci, "A model of eye movements and visual attention," in *Proceedings of the International Conference on Cognitive Modeling*, 2000, pp. 252–259.

[41] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[42] P. D. Sherman, *Colour vision in the nineteenth century: the Young-Helmholtz-Maxwell theory*. Taylor & Francis, 1981.

[43] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," *Vision Research*, vol. 45, no. 5, pp. 643–659, 2005.

[44] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[45] P.-H. Tseng, R. Carmi, I. G. Cameron, D. P. Munoz, and L. Itti, "Quantifying center bias of observers in free viewing of dynamic natural scenes," *Journal of Vision*, vol. 9, no. 7, p. 4, 2009.

[46] M. Van den Bergh, X. Boix, G. Roig, B. de Capitani, and L. Van Gool, "SEEDS: Superpixels extracted via energy-driven sampling," in *European Conference on Computer Vision*. Springer, 2012, pp. 13–26.

[47] M. E. Wieser, "Atomic weights of the elements 2005," *Pure and Applied Chemistry*, vol. 78, pp. 2051–2066, 2006.

[48] W. D. Wright, “A re-determination of the trichromatic coefficients of the spectral colours,” *Transactions of the Optical Society*, vol. 30, no. 4, p. 141, 1929.

[49] S. Zeki, J. Watson, C. Lueck, K. J. Friston, C. Kennard, and R. Frackowiak, “A direct demonstration of functional specialization in human visual cortex,” *The Journal of Neuroscience*, vol. 11, no. 3, pp. 641–649, 1991.

[50] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “SUN: A Bayesian framework for saliency using natural statistics,” *Journal of Vision*, vol. 8, no. 7, p. 32, 2008.

Appendix A
Smoothed Saliency Map Experiment

It was observed that the click and eye-tracking data were quite random and haphazardly spread across the images, whereas the saliency maps have pixel-level accuracy and preserve object boundaries quite well. It was questioned as to what effect accommodating the randomness in the data by smoothing the saliency maps would have on the saliency scores. To test this, the exact same experiment as detailed in Chapter 5 is run except for smoothing the saliency maps with a Gaussian filter with $\sigma = 0.5°$ of visual angle, or 9px [28], before computing the scores.

The smoothed equi-width, equi-depth and diagonal histogram results tables are found in Table A.1, Table A.2 and Table A.3 respectively, with the percentage increase in using the smoothed saliency maps with equi-depth histograms over the equi-width histograms found in Table A.4. In addition, a percentage increase in using the smoothed saliency maps over the unsmoothed saliency maps is presented in Table A.5. As can be seen there are significant improvements in scores, with only the sAUC metric being relatively invariant to the smoothing. This again highlights the sensitivity of these scoring methods and calls for more robust methods to be designed.

Table A.1: Equi-width histogram results using smoothed saliency maps (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.506	0.454	0.308	0.437	0.502	0.457	0.308	0.431	**0.399**	0.372	0.250	0.326	0.457	0.410	0.286	0.411
	0.512	0.458	0.292	0.428	0.502	0.454	0.288	0.416	0.376	0.347	0.217	0.298	0.479	0.428	0.284	0.415
	0.442	0.393	0.260	0.392	0.443	0.399	0.263	0.389	0.369	0.343	0.227	0.307	0.394	0.349	0.238	0.365
	0.514	0.458	0.286	0.425	**0.503**	0.452	0.281	0.412	0.376	0.343	0.210	0.291	**0.480**	0.428	0.279	0.412
NSS	**0.817**	0.737	0.497	0.701	**0.820**	0.751	0.511	0.699	**1.081**	1.037	0.765	0.857	0.754	0.676	0.479	0.679
	0.816	0.731	0.461	0.676	0.809	0.736	0.464	0.664	0.949	0.901	0.595	0.711	**0.781**	0.698	0.467	0.677
	0.706	0.629	0.412	0.621	0.720	0.654	0.434	0.629	1.029	0.985	0.718	0.831	0.635	0.560	0.388	0.590
	0.816	0.727	0.447	0.668	0.805	0.728	0.448	0.653	0.930	0.864	0.567	0.677	0.778	0.694	0.453	0.669
AUC	**0.725**	0.703	0.639	0.697	**0.724**	0.705	0.640	0.695	**0.745**	0.730	0.668	0.709	0.711	0.690	0.636	0.691
	0.723	0.698	0.624	0.688	0.720	0.699	0.624	0.685	0.730	0.715	0.642	0.689	0.712	0.689	0.628	0.687
	0.693	0.670	0.614	0.671	0.695	0.675	0.616	0.673	0.722	0.709	0.651	0.695	0.677	0.655	0.610	0.665
	0.724	0.699	0.620	0.688	0.722	0.699	0.620	0.684	0.727	0.707	0.635	0.681	**0.713**	0.690	0.624	0.686
sAUC	**0.548**	0.546	0.528	0.530	**0.557**	0.556	0.537	0.537	**0.619**	0.619	0.593	0.588	**0.534**	0.532	0.525	0.522
	0.532	0.527	0.512	0.514	0.538	0.535	0.519	0.519	0.590	0.590	0.562	0.557	0.520	0.517	0.512	0.509
	0.532	0.529	0.515	0.516	0.542	0.541	0.525	0.526	0.605	0.607	0.582	0.580	0.518	0.516	0.511	0.511
	0.526	0.522	0.505	0.510	0.533	0.531	0.512	0.514	0.582	0.577	0.553	0.544	0.515	0.512	0.506	0.505

Table A.2: Equi-depth histogram results using smoothed saliency maps (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.545	0.508	0.304	0.492	**0.535**	0.503	0.302	0.477	**0.403**	0.383	0.238	0.340	0.505	0.471	0.286	0.474
	0.556	0.520	0.289	0.484	0.528	0.496	0.276	0.454	0.350	0.330	0.181	0.284	**0.546**	0.512	0.293	0.488
	0.507	0.470	0.274	0.464	0.499	0.466	0.274	0.451	0.380	0.359	0.220	0.324	0.467	0.432	0.255	0.444
	0.555	0.502	0.283	0.469	0.531	0.476	0.273	0.437	0.363	0.310	0.188	0.268	0.541	0.497	0.283	0.474
NSS	0.871	0.812	0.484	0.779	**0.862**	0.813	0.495	0.758	**0.995**	0.959	0.680	0.776	0.830	0.773	0.474	0.777
	0.874	0.816	0.448	0.753	0.833	0.784	0.433	0.704	0.757	0.720	0.429	0.549	**0.891**	0.834	0.479	0.795
	0.803	0.742	0.428	0.727	0.798	0.747	0.443	0.712	0.946	0.909	0.639	0.748	0.753	0.695	0.413	0.717
	0.874	0.781	0.439	0.724	0.840	0.744	0.431	0.672	0.816	0.646	0.474	0.493	0.880	0.801	0.462	0.766
AUC	0.742	0.726	0.636	0.719	**0.739**	0.725	0.636	0.715	**0.746**	0.735	0.659	0.711	0.730	0.715	0.636	0.716
	0.745	0.729	0.626	0.713	0.734	0.721	0.620	0.701	0.711	0.699	0.610	0.665	**0.741**	0.727	0.636	0.718
	0.722	0.704	0.619	0.704	0.720	0.704	0.622	0.700	0.726	0.716	0.642	0.699	0.708	0.691	0.617	0.699
	0.744	0.720	0.621	0.705	0.735	0.711	0.617	0.692	0.720	0.681	0.616	0.649	0.738	0.718	0.629	0.711
sAUC	**0.538**	0.534	0.523	0.516	**0.546**	0.544	0.532	0.523	**0.597**	0.595	0.580	0.554	**0.526**	0.524	0.522	0.514
	0.504	0.501	0.497	0.484	0.506	0.505	0.499	0.483	0.522	0.522	0.517	0.473	0.503	0.501	0.503	0.492
	0.530	0.526	0.512	0.510	0.537	0.535	0.523	0.516	0.581	0.582	0.567	0.548	0.517	0.515	0.511	0.507
	0.506	0.490	0.496	0.477	0.509	0.492	0.501	0.473	0.538	0.499	0.528	0.455	0.502	0.492	0.501	0.487

Table A.3: Diagonal histogram results using smoothed saliency maps (bold and underlined indicates highest score for combination of global contrast and spatial sparsity per scoring measure and dataset task).

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	0.532	0.490	0.307	0.470	**0.524**	0.488	0.305	0.460	**0.406**	0.385	0.247	0.339	0.488	0.449	0.285	0.448
	0.540	0.498	0.288	0.459	0.518	0.482	0.278	0.437	0.360	0.339	0.192	0.289	**0.522**	0.483	0.288	0.456
	0.471	0.431	0.267	0.431	0.468	0.434	0.269	0.425	0.376	0.355	0.230	0.324	0.426	0.389	0.244	0.406
	0.537	0.493	0.281	0.453	0.516	0.476	0.271	0.431	0.362	0.334	0.190	0.284	0.517	0.476	0.280	0.450
NSS	**0.855**	0.789	0.492	0.749	**0.852**	0.797	0.506	0.738	**1.070**	1.039	0.748	0.844	0.802	0.739	0.475	0.736
	0.853	0.787	0.449	0.718	0.824	0.769	0.440	0.685	0.843	0.810	0.494	0.620	**0.851**	0.786	0.472	0.744
	0.752	0.688	0.423	0.680	0.759	0.706	0.446	0.680	1.021	0.992	0.728	0.835	0.690	0.629	0.397	0.659
	0.847	0.774	0.435	0.705	0.818	0.755	0.426	0.671	0.840	0.779	0.485	0.595	0.838	0.771	0.453	0.729
AUC	0.737	0.719	0.637	0.709	**0.734**	0.719	0.637	0.705	**0.749**	0.737	0.664	0.713	0.722	0.706	0.635	0.705
	0.737	0.720	0.623	0.703	0.729	0.714	0.619	0.694	0.720	0.709	0.621	0.675	**0.731**	0.714	0.632	0.706
	0.707	0.688	0.616	0.690	0.707	0.691	0.620	0.689	0.726	0.716	0.652	0.702	0.693	0.675	0.614	0.684
	0.736	0.716	0.618	0.700	0.728	0.710	0.615	0.691	0.720	0.700	0.617	0.667	0.729	0.711	0.626	0.702
sAUC	**0.541**	0.538	0.526	0.521	**0.550**	0.549	0.535	0.527	**0.611**	0.608	0.589	0.573	**0.528**	0.526	0.524	0.517
	0.510	0.510	0.499	0.496	0.514	0.514	0.503	0.497	0.552	0.554	0.534	0.512	0.507	0.505	0.505	0.499
	0.532	0.529	0.518	0.517	0.541	0.540	0.529	0.525	0.599	0.598	0.583	0.573	0.520	0.517	0.516	0.513
	0.508	0.503	0.495	0.492	0.512	0.509	0.501	0.493	0.550	0.540	0.530	0.501	0.504	0.501	0.501	0.496

Table A.4: Percentage increase (or decrease) in scores when using smoothed saliency maps generated with equi-depth histograms over equi-width histograms. Green indicates that using equi-depth histograms improved scores and red indicates that they worsened scores.

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	8%	12%	-1%	13%	6%	10%	-2%	11%	1%	3%	-5%	4%	10%	15%	0%	15%
	9%	14%	-1%	13%	5%	9%	-4%	9%	-7%	-5%	-16%	-5%	14%	20%	3%	18%
	15%	20%	5%	19%	13%	17%	4%	16%	3%	5%	-3%	6%	18%	24%	7%	22%
	8%	10%	-1%	10%	6%	5%	-3%	6%	-3%	-9%	-10%	-8%	13%	16%	2%	15%
NSS	7%	10%	-2%	11%	5%	8%	-3%	8%	-8%	-8%	-11%	-9%	10%	14%	-1%	14%
	7%	12%	-3%	11%	3%	6%	-7%	6%	-20%	-20%	-28%	-23%	14%	20%	3%	17%
	14%	18%	4%	17%	11%	14%	2%	13%	-8%	-8%	-11%	-10%	19%	24%	6%	22%
	7%	7%	-2%	8%	4%	2%	-4%	3%	-12%	-25%	-16%	-27%	13%	15%	2%	14%
AUC	2%	3%	0%	3%	2%	3%	-1%	3%	0%	1%	-1%	0%	3%	4%	0%	4%
	3%	5%	0%	4%	2%	3%	-1%	2%	-3%	-2%	-5%	-4%	4%	6%	1%	4%
	4%	5%	1%	5%	4%	4%	1%	4%	1%	1%	-1%	1%	5%	6%	1%	5%
	3%	3%	0%	3%	2%	2%	0%	1%	-1%	-4%	-3%	-5%	3%	4%	1%	4%
sAUC	-2%	-2%	-1%	-3%	-2%	-2%	-1%	-3%	-4%	-4%	-2%	-6%	-2%	-2%	-1%	-2%
	-5%	-5%	-3%	-6%	-6%	-6%	-4%	-7%	-11%	-12%	-8%	-15%	-3%	-3%	-2%	-3%
	0%	-1%	-1%	-1%	-1%	-1%	0%	-2%	-4%	-4%	-3%	-5%	0%	0%	0%	-1%
	-4%	-6%	-2%	-6%	-5%	-7%	-2%	-8%	-7%	-13%	-5%	-16%	-3%	-4%	-1%	-4%

Table A.5: Percentage increase (or decrease) in the equi-width scores using the smoothed saliency maps over the unsmoothed saliency maps.

	Free View Task				Saliency Left/Right Task				Explicit Click Task				Object Search Task			
CC	15.0%	18.9%	30.3%	17.4%	14.8%	18.5%	29.9%	17.2%	13.9%	17.1%	27.1%	16.7%	15.2%	19.1%	30.4%	17.1%
	14.0%	18.3%	32.3%	17.3%	13.9%	18.1%	32.3%	17.2%	13.7%	17.7%	32.0%	17.6%	13.7%	17.9%	30.9%	16.6%
	20.4%	24.2%	36.6%	21.0%	20.1%	23.6%	35.9%	20.8%	18.5%	21.3%	31.2%	19.6%	20.7%	24.8%	37.2%	20.8%
	12.5%	17.1%	32.1%	16.6%	12.5%	16.9%	32.1%	16.6%	12.3%	16.8%	32.2%	17.3%	12.3%	16.7%	30.5%	15.9%
NSS	14.4%	18.2%	29.8%	17.4%	12.9%	16.2%	25.6%	15.7%	2.4%	4.1%	6.7%	7.3%	14.3%	18.2%	29.3%	16.4%
	13.8%	18.1%	32.8%	17.6%	12.6%	16.4%	29.7%	16.3%	4.3%	6.5%	15.7%	12.6%	12.8%	17.0%	30.3%	16.1%
	20.1%	23.9%	36.8%	21.4%	18.1%	21.2%	31.2%	19.2%	6.1%	7.3%	9.8%	9.4%	20.6%	24.6%	37.4%	20.7%
	12.4%	16.9%	33.1%	17.0%	11.3%	15.4%	29.8%	15.9%	4.3%	7.6%	16.9%	14.4%	11.7%	15.9%	30.7%	15.6%
AUC	4.8%	5.5%	6.3%	5.1%	4.7%	5.7%	6.4%	4.8%	3.5%	4.2%	5.1%	4.1%	4.6%	5.5%	6.1%	4.7%
	5.4%	6.2%	6.2%	5.0%	5.0%	6.0%	6.1%	4.8%	4.1%	5.3%	5.4%	4.3%	4.9%	5.8%	6.0%	4.9%
	6.2%	6.7%	7.0%	5.6%	6.4%	6.7%	7.0%	5.5%	5.0%	6.0%	6.2%	4.5%	6.3%	6.5%	7.1%	5.4%
	4.8%	6.0%	5.9%	4.8%	4.9%	6.0%	5.8%	4.7%	4.3%	5.5%	5.2%	4.3%	4.6%	5.5%	5.9%	4.4%
sAUC	1.2%	1.5%	1.3%	1.4%	1.4%	1.9%	1.6%	1.1%	0.8%	1.3%	1.7%	1.4%	0.8%	1.2%	1.0%	0.7%
	1.4%	1.3%	0.3%	0.5%	1.1%	1.2%	0.7%	0.3%	1.2%	1.7%	1.2%	1.1%	0.4%	0.6%	0.2%	0.1%
	1.5%	1.7%	1.2%	1.2%	2.0%	1.9%	1.7%	1.1%	1.6%	2.5%	2.6%	1.3%	1.3%	1.3%	1.2%	0.8%
	0.6%	1.0%	0.0%	0.4%	1.0%	1.4%	0.3%	0.3%	1.3%	1.8%	0.9%	0.9%	0.3%	0.4%	0.4%	-0.1%