REM: Relational Entropy-Based Measure of Saliency

Kester Duncan*
Computer Science and Engineering
University of South Florida
Tampa, FL 33620, U.S.A.
kkduncan@cse.usf.edu

Sudeep Sarkar
Computer Science and Engineering
University of South Florida
Tampa, FL 33620, U.S.A.
sarkar@cse.usf.edu

ABSTRACT

Human eye fixation points occurring during the early stages of visual processing often correspond to the loci of salient image regions. These salient regions provide us with assistance in determining the interesting parts of an image and they also lend support to our ability to discriminate between different objects in a scene. They attract our immediate attention without requiring an exhaustive scan of a scene and they possess some quality that enables them to stand out in relation to their neighbors. In this paper, we present a bottom-up measure of saliency based on the relationships exhibited among image features. We adopt the standpoint whereby the relationships among features determines more of the perceived structure in an image rather than the individual feature attributes and we seek those structures which 'pop-out.' We capture the organization within an image by employing relational distributions derived from distance and gradient direction relationships exhibited between image pixels. We demonstrate how our results coincide with human fixations. We also evaluate the performance of our measure in relation to a dominant saliency model and obtain comparable results. In an effort to derive meaningful information from an image, we investigate the significance of scale relative to our saliency measure.

Keywords

Bottom-Up, Rényi Entropy, Relational Histograms, Scale-Variation, Saliency

1. INTRODUCTION

Certain structures or regions in a scene often attract our immediate attention without requiring an exhaustive scan of the scene itself. The way these regions are captured by the Human Visual System (HVS) without the need for focused attention is often described as *pre-attentive* processing which was suggested by Neisser in [22] as the first of the two stages of human visual processing. This stage consists of parallel processes that operate concurrently on large regions of the visual field, forming structures to which attention can be directed. Anything perceived within the pre-attentive time frame (which is typically 200 milliseconds) incorporates only the information available from a single cursory glimpse [13]. An immediate visual arousal occurs in the early stages of human visual processing [17] as a result of the pre-attentively distinctive parts of a scene, and it is this idea that is commonly referred to as *saliency*. With regards to computer vision, saliency can be defined as the quality of an image feature that allows it to stand out in relation to its neighboring features. These features are almost unique, thereby making it possible to discriminate between objects in a scene. It must be noted however, that salient regions in an image may not necessarily belong to an object of interest.

Saliency can often provide the foundation for a visual attention mechanism whereby the need for computational resources is significantly reduced [12]. The selection of a commensurate set of salient features forms the first step in many computer vision algorithms. Salient features, points, or regions, facilitate object recognition, perceptual organization, segmentation, and figure-ground separation because they permit immediate concentration on objects of interest in an image.

Consequently, most saliency mechanisms are commonly approached in two ways: *bottom-up* saliency and *top-down* saliency. Bottom-up approaches are analogous to rapid, image or stimulus-driven mechanisms in pre-attentive vision and are to a great extent independent of the knowledge of the content in an image. In [25], Sha'ashua and Ullman presented a saliency measure based on curvature and curvature variation. The structures their measure emphasized were also salient in human perception, often corresponding to objects of interest in the image. The authors proposed a bottom-up mechanism for detecting salient locations using a locally connected network.

Similarly, Kadir and Brady introduced a multi-scale algorithm for salient region selection [17]. Their technique determined salient regions as those exhibiting unpredictable characteristics simultaneously in some feature-space and over scale. They used the local intensity as the descriptor for saliency. In this paper, we also investigate the implications of scale and saliency and adopt the argument presented in [17] that scale "is intimately related to the issue of determining saliency and extracting relevant descriptions." We explore this in our measure to determine the scale at which a pixel remains salient.

*Corresponding author

ICVGIP '10, December 12-15, 2010, Chennai, India

Motivated by the work done in [25], Berengolts and Lindenbaum [2] presented a saliency measure based on probabilistic cues, estimated length distributions and the expected length of curves. They demonstrated this using a process based on gray-level similarity. Additionally, Hou and Zhang presented a saliency detection mechanism based on the log spectra representation of images [14]. More recently, Avraham and Lindenbaum proposed a novel stochastic model to estimate saliency in [4]. Their "esaliency" mechanism determines if an image part is of interest with the goal of finding small image regions where salient objects are present. In this paper, we present a purely bottom-up, task independent measure for saliency detection.

Conversely, top-down approaches are goal-oriented and utilize prior knowledge about the scene or the context to identify salient regions [11]. They are task-dependent thereby demanding a more thorough understanding of the context of the image. For example, Gao, Han, and Vasconcelos couple saliency to the object recognition goal in [8]. They argued that the saliency judgments become significantly more adaptive, only highlighting image areas which were relevant to recognition. The authors equated saliency to discrimination and they referred to the optimal salient features as those that were maximally informative of the presence or absence of the target class in a field of view [8]. Gao and Vasconcelos proceeded to define discriminant saliency in [9] as the notion whereby the features whose response best distinguishes an object to be recognized from the set of all others that may be of possible interest. Additionally, Gopalakrishnan, Hu, and Rajan presented a salient region detection framework based on the color and orientation distribution in images [11]. This framework consisted of a color saliency framework and an orientation framework. Rather than analyzing localized features, they considered the global distribution of color and different orientations.

Bottom-up and top-down approaches together are analogous to pre-attentive and attentive vision respectively. Integration of these two approaches has been deemed crucial for robot navigation, visual surveillance, and realistic visual searches [21]. Saliency mechanisms utilizing this approach fall into another category known as *integrated* approaches. Itti and Navalpakkam united bottom-up and top-down approaches of saliency for a novel approach in [21]. Their method decomposed the visual input into a set of topographic feature maps. The bottom-up component was responsible for computing the saliency of locations in different feature maps whereas the top-down component used statistical knowledge of the target object to tune the bottom-up maps. Similarly, Goferman et al. utilized local, global, and high-level factors for their context-aware saliency detection mechanism in [10] in an effort to detect the important parts of a scene. Salient regions detected by their approach contained the prominent objects as well as parts of the background that conveyed the context.

Section 2, we present our relational entropy-based measure of saliency first by describing relational distributions, then we move into a brief description of Rényi entropy. In Section 3, we present our results on various images, including saliency maps and analyses. We compare our findings to those produced by iLab's [20, 28] saliency model. In Section 4 we elaborate on our findings and discuss the implications of our work on saliency research.

2. THEORY

Our saliency measure is formulated on the entropy of geometric relational distributions. These topics are described in detail in the subsequent sections.

2.1 Relational Distributions

We adopt the notion specified in [27] that "the structure perceived in an image is determined more by the relationships among image features rather than by the individual feature attributes." We utilize a mechanism to capture this structure. Image structures can be represented by probability functions referred to as relational distributions. We capture these distributions using relational histograms. The concept of relational histograms is not a novel one and they have been used particularly for database indexing [15], motion-based recognition of humans [27], shape analysis [23], and object recognition [3]. We formally define relational distributions in definitions 1 and 2 following [27].

Definition 1. Let:

- $\mathcal{F} = \{f_1, ..., f_N\}$ represent the set of N features in an image.
- $\mathbf{F_k}$ represent a random k-tuple of features, and
- The relationship among these k-tuple features be denoted by R_k.

Therefore, pairwise or binary relationships between features are represented by R_2. Low-order spatial dependencies are captured by small values of k whereas higher-order dependencies are captured by larger values of k.

Definition 2. Let the relationships R_k be characterized by a set of M attributes $\mathbf{A_k} = \{A_{k1}, ..., A_{kM}\}$. Hence, image structures can be represented by joint probability functions: $P(\mathbf{A_k} = \mathbf{a_k})$, also denoted by $P(a_{k1}, ..., a_{kM})$ or $P(\mathbf{a_k})$, where a_{ki} is the value taken by the relational attribute A_{ki}.

These distributions can be interpreted as: Given an image, if you pick k-tuples of features, what is the probability that it will exhibit the relational attributes $\mathbf{a_k}$ or $P(\mathbf{A_k} = \mathbf{a_k})$?

2.1.1 Pixel-based features

The concept of relational distributions is illustrated by considering the pixel properties as features. Each pixel, f_i, is associated with the gradient direction, θ_i, estimated using a Canny Edge detector. To capture some structure between two pixels, we use the difference between gradient angles $(\theta_i - \theta_j)$ and the euclidean distance $(d_i - d_j)$ between them as the attributes, $\{A_{21}, A_{22}\}$, of R_2. These attributes are ideal because they are invariant with respect to image plane rotation and translation. Figure 1 depicts the computed attributes. In addition to these attributes, we also utilize the gradient magnitude differences between pixels as weights w_i for histogram bin voting.

2.2 Saliency Measure

Let $\mathcal{P} = (p_1, p_2, ..., p_n)$ be a discrete probability distribution. The amount of uncertainty or disorder of the distribution $\mathcal{P}$ is referred to as the entropy of $\mathcal{P}$ and it is measured by the quantity $H[\mathcal{P}] = H(p_1, p_2, ..., p_n)$ [24]. In our case, the distribution is $\mathcal{P}(d, \theta)$. Entropy designates the extent to

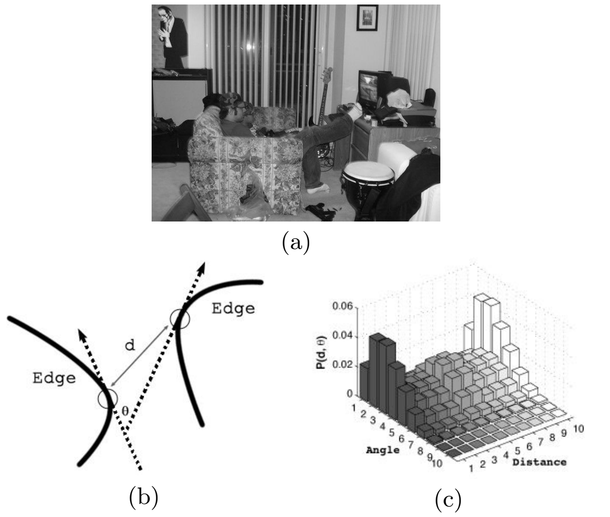


Figure 1: Pixel-based binary relational distribution. (a) Original Image, (b) The two attributes that characterize the relationship between two pixels - distance and gradient angle. (c) 3D illustration of the relational distribution $P(d, \theta)$ formed from the image in (a).

which the features characterized by the relational distribution are uniformly distributed [5]. Entropy is defined by the common form as:

$$H(P) = -\sum_{i=1}^{n} p(x_i)\ \log_2\ p(x_i), \tag{1}$$

which is universally known as *Shannon's Entropy.* We utilize the Rényi entropy which is a generalization of the Shannon entropy given in Equation 1. It is defined as follows:

$$H_\alpha(P) = \frac{1}{1-\alpha}\ \log_2\ (\ \sum_{i=1}^{n} p(x_i)^\alpha\ ) \tag{2}$$

Equation 2 is known as *Rényi's Entropy* of order α where $\alpha \geq 0$. Increasing values of α produce a Rényi entropy that is devised by favoring the higher probability events. The probability events are considered more equally for lower values of α. When $\alpha = 1$, we get the Shannon entropy. In this work, an α value of 2 was utilized: This is referred to as the *Collision entropy.* Furthermore, we utilize an extra term $(l-1)\frac{log_2(e)}{2N}$ shown by Abe in [1] to be the "the expected divergence between a finite probability distribution Q on $\{1, 2, ..., l\}$ and its empirical one obtained from the sample of size N drawn from Q". This term is added to the entropy value $H[P]$ in an effort to compute the expected divergence between the estimated probabilities and the actual underlying probability as shown in Equation 3.

$$H_2(P) = -\log_2\ (\ \sum_{i=1}^{n} p(x_i)^2\ ) + (l-1)\frac{\log_2(e)}{2N} \tag{3}$$

We believe that finding the entropy of the relational distribution $\mathcal{P}(d, \theta)$ is a good indicator of the "pop-out" structures in an image.

Furthermore, we define saliency as the quality of an image feature that enables it to stand out (or precisely "pop-out") relative to its neighbors. We quantify this quality by calculating the Rényi entropy of the relational distributions of local pixel neighborhoods. Thus, the saliency function $\Phi(\cdot)$ is defined as:

$$\Phi = 1 - H_2[\mathcal{P}(d, \theta)] \tag{4}$$

Φ is a measure of pixel saliency with regards to M, where M is a $(2k+1)(2k+1)$ neighborhood of a pixel with $k > 1$. Higher values of Φ indicate greater saliency and vice versa. We consider the pairwise comparisons of pixels in the neighborhood M_i of a central pixel f_i, where i is the index of the pixel. Examples of neighborhoods used are shown in Figure 2.

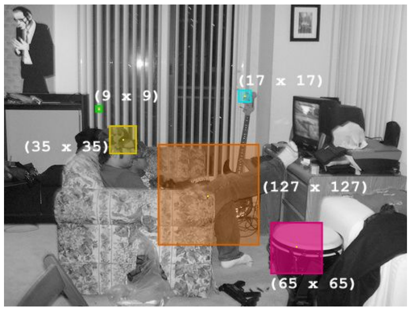


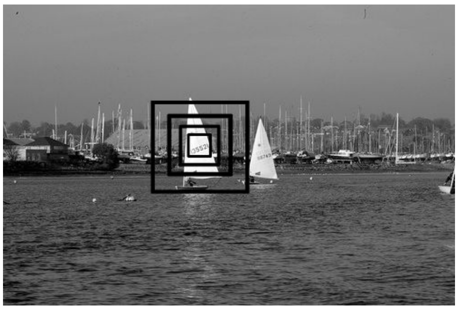
Figure 2: Illustration of some of the local pixel neighborhoods that are used to measure the saliency of a central pixel (depicted as a yellow dot).

2.3 Scale Variation

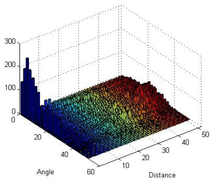
Objects in the world appear in different ways depending on the scale of observation and this fact has important implications if they are to be accurately described. Multi-scale representations are necessary to completely represent and process images [26]. A characteristic property of structures in images is that they may only be meaningful over definite ranges of scale. For instance, a map of the United States would contain the largest cities, towns, and some interstate highways, whereas a city map changes the level of abstraction substantially to include streets and buildings etc.

The images in Figure 3 bolster our approach. They depict the relational distributions of varying local pixel neighborhood sizes ranging from (17 x 17) to (129 x 129) pixels for a central pixel located in the sail of the sailboat (the respective neighborhoods are outlined in the image in Figure 3(a)). The horizontal and vertical axes for each graph represent the pairwise pixel distances and gradient angle differences respectively. From the image in Figure 3 (a), we see that the sailboats are the most conspicuous objects, therefore they can be considered the most salient image regions. We can see that as the scale increases, the amount of information available in the distribution increases and the relationships between the pixels change leading to changes in entropy (Figure 3(f)). However, after the (33 x 33) scale, the changes are relatively negligible.

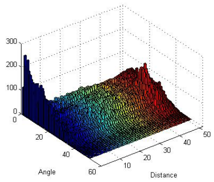
In computer vision, the primary focus is on deriving significant and meaningful information from images. In accordance with this notion, we explore the significance of scale relative to our saliency measure and attempt to elicit the optimal scales for their analysis [18]. The optimal scale in our case is the scale up to which the high saliency of an



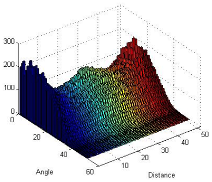
(a) Test image



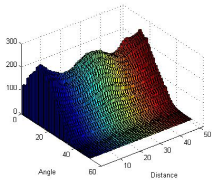
(b) 17 x 17



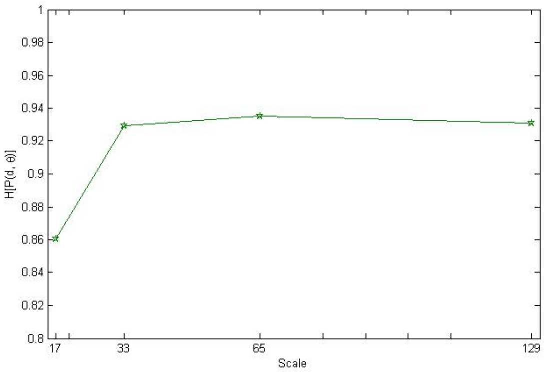
(c) 33 x 33



(d) 65 x 65



(e) 129 x 129



(f) Entropy Plot

Figure 3: Relational Distributions of different pixel neighborhood scales (this figure is best viewed in color).

image point persists or stabilizes. We do not attempt to select the optimal scales automatically. Representations of scale-variation would enable us to analyze an image point of interest at different scales, however they do not indicate at which scale subsequent processing should be performed. This is the subject of future work. Our saliency map emphasizes salient locations in an image at a specified scale. We process an image at different local pixel neighborhood scales for square neighborhood sizes satisfying $2^k + 1$ dimensions, where $k = 1, ..., 5$.

3. RESULTS AND ANALYSES

We present analyses and evaluations of our saliency measure in this section. We investigated its performance for a wide variety of images and compared our results with a dominant saliency model. As previously noted, our measure is a pure bottom-up, task-independent approach to saliency detection. There is no knowledge about the context of the scene that is used to determine saliency. Salient regions are simply those regions which stand out relative to their neighborhood. Since we utilize gradient information, namely gradient direction and magnitude, boundaries of salient regions are emphasized rather than their interior. This is due to the fact that within the salient region (if composed of many salient pixels), there may be nothing that stands out locally, hence uniformity. This can be observed in the subsequent sections.

3.1 Comparison with human saliency maps

We analyze the performance of our measure in relation to that obtained by the human visual attention mechanism. We executed this by comparing our saliency maps with empirical human fixation maps (or fixation density maps) from work done in [6]. These human saliency maps were captured by recording human eye fixations over an image which was displayed to test subjects for a limited amount of time. Each fixation point in these images were then convolved with a Gaussian filter. We subjectively compare our results in Figure 4 to provide an approximate evaluation of the correlation between the human fixation map and the REM saliency map visually.

3.2 Comparison with a dominant saliency model

To compare our results objectively, we utilize the correlation coefficient between the human fixation map and our REM saliency map. The correlation coefficient λ is calculated as follows:

$$\lambda = \frac{\sum_x[(M_h(x) - \mu_h) \cdot (M_s(x) - \mu_s)]}{\sqrt[2]{\sum_x (M_h(x) - \mu_h)^2 \cdot \sum_x (M_s(x) - \mu_s)^2}} \quad (5)$$

where $M_h(x)$ is the human fixation map, $M_s(x)$ is the REM saliency map, μ_h is the mean intensity of the human fixation map ($M_c(x)$), and μ_s is the mean intensity of our map ($M_s(x)$). Figure 4(e) shows the correlation coefficients between the human fixation maps and the REM saliency maps for the respective images.

We compared the results of our algorithm to that of the results produced by iLab's [19, 28] (available in [20]) saliency mechanism in relation to the respective human saliency maps. iLab's saliency mechanism is considered to be the dominant saliency model in the state of the art. The REM saliency maps were produced by evaluating a local pixel neighborhood size of 11x11. We used 120 human fixation maps from the Bruce and Tsotsos data set (some examples are shown in Figure 4). The overall performance for these 120 images is displayed in Figure 5. In 63.3% of the images, the correlation coefficients between the R.E.M. saliency maps and the human fixation maps were higher than iLab's, hence showing that our method produces results which are more consistent with the HVS.

3.3 Scale Variation results

In this section, we evaluate the changes that occur to a saliency map over a narrow range of scales. The scale range

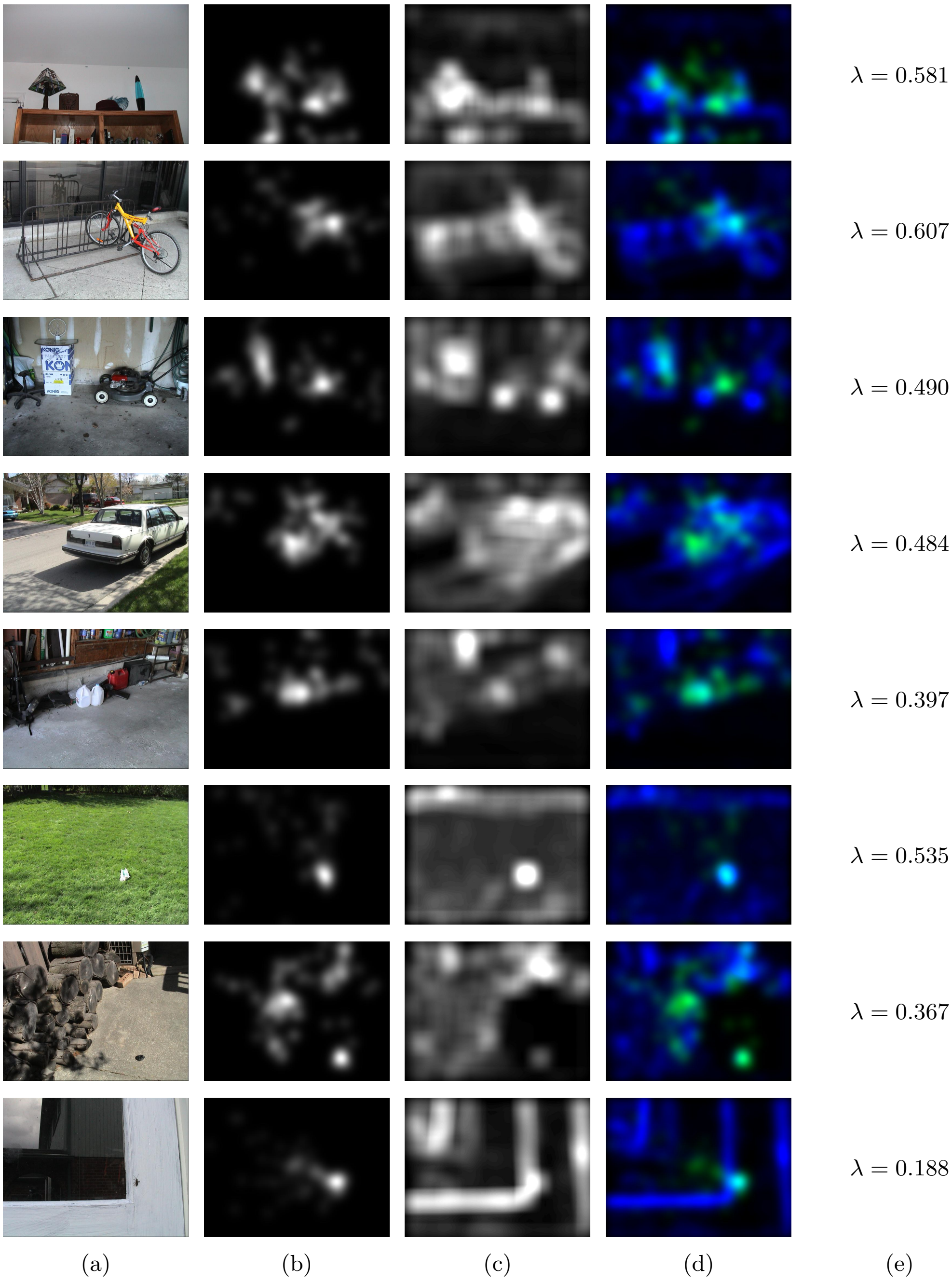


Figure 4: Comparing saliency results and *human saliency maps* for images taken from the Bruce and Tsotsos dataset [6]. (a) Original image. (b) Human Saliency map. (c) *REM* saliency map. (d) *Comparison map* - a different color channel is assigned to each map - blue for our saliency map, green for the human fixation map, and red 0. (e) Correlation coefficients λ between human saliency maps and REM saliency maps.

(σ) is as follows: $(2^k + 1)$ x $(2^k + 1)$, where $k = \{1, ..., 5\}$; $\sigma = n$ would always refer to (n x n). The saliency maps were intensity-normalized and smoothed with a Gaussian smoothing filter.

From the images in Figure 6, we can subjectively conclude that the border of the ceiling lights and the helmet reflections are the most salient over these narrow range of scales. All other image regions fade to the 'background' as the scale increases. For the images in Figure 7, we can conclude that the borders of the chairs are the most salient over these narrow range of scales. The heads of the band members persist up to the $\sigma = 17$ scale, but they are not so apparent at $\sigma = 33$. Correspondingly, in Figure 8, the borders of the individual's t-shirt persist up to the $\sigma = 17$ scale thereby suggesting that they are more salient that other image regions. Moreover, at the largest scale, most image structures are

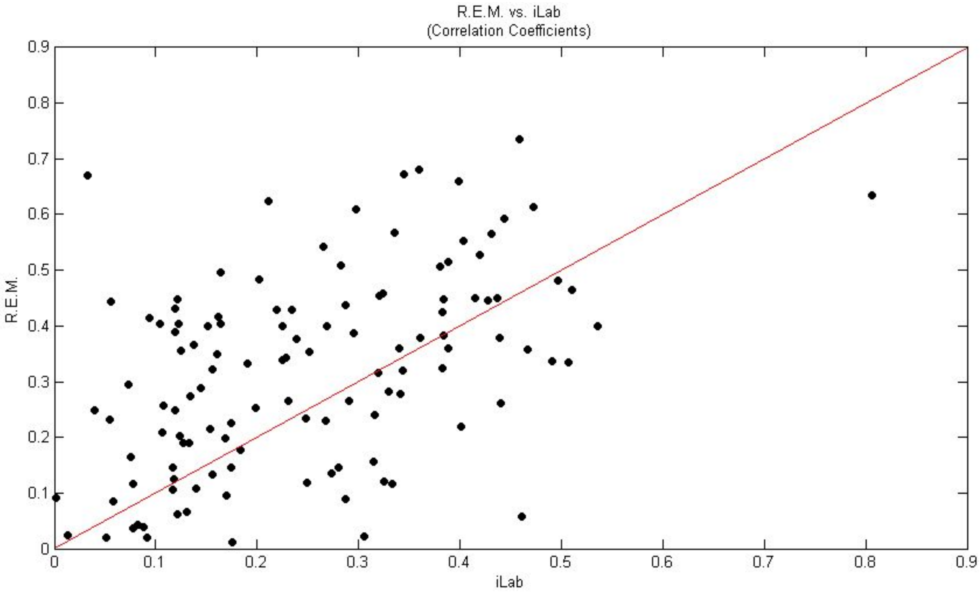


Figure 5: R.E.M. vs. iLab - graph of the correlation coefficients for 120 images from the Bruce and Tsotsos data set. In 63.3% of the images, the correlation coefficients between the R.E.M. saliency maps and the human fixation maps were higher than iLab's.

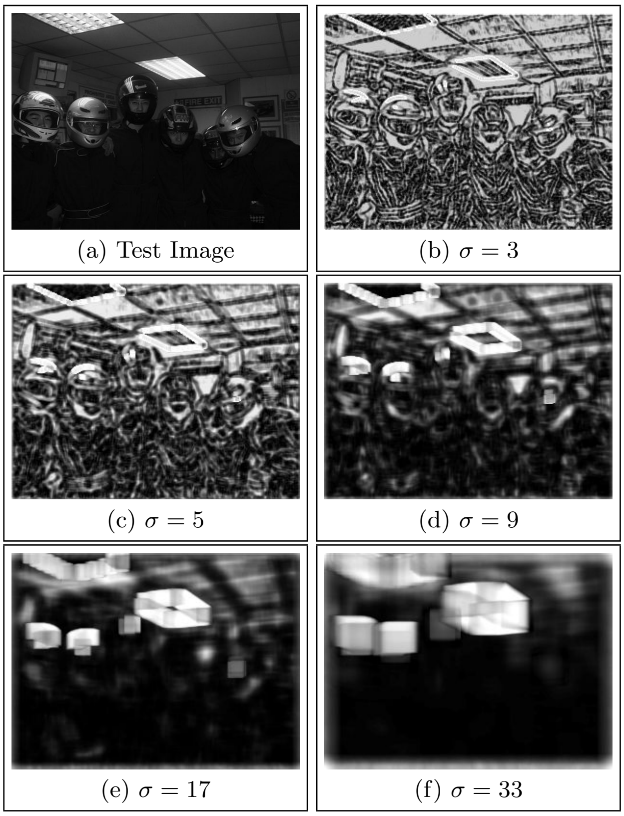


Figure 6: Scale variation of the *helmets* image. (Image taken from the PASCAL dataset [7].)

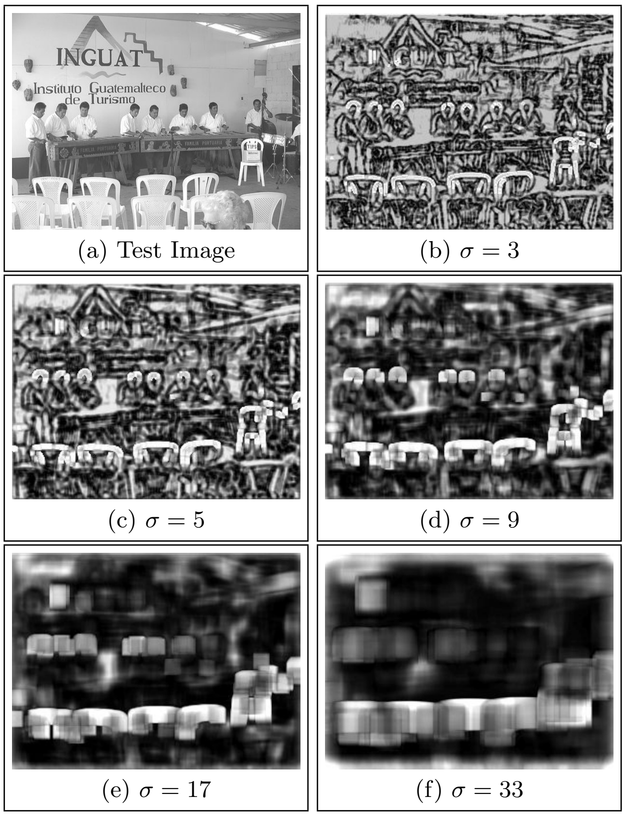


Figure 7: Scale variation of the *band* image. (Image taken from the PASCAL dataset [7].)

blurry, suggesting that they do not possess strong saliency at this scale. We can see from these images that the most

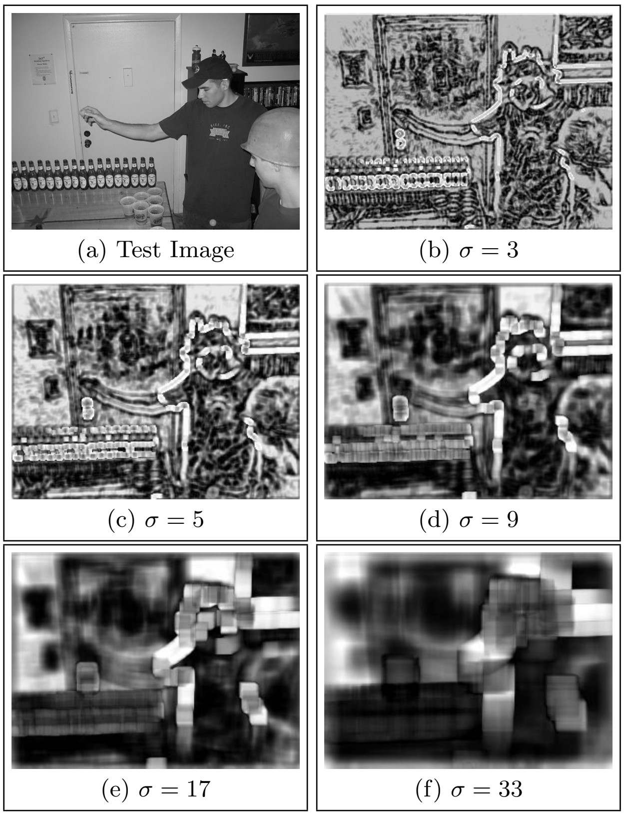


Figure 8: Scale variation of the *beer_bottles* image. (Image taken from the PASCAL dataset [7].)

salient image regions persist through to the largest scale. Less salient regions fade as the scale increases. For future work, we aim to automatically select the appropriate scale of observation for each image point following work done in [17].

4. CONCLUSION AND FUTURE WORK

The main goal of this paper was to develop a pure bottom-up saliency mechanism based on relationships exhibited between image features. We highlighted those image regions which stood out relative to some local pixel neighborhood. We adopted a bottom-up saliency approach due to its generic nature and flexibility. Our measure is not tied to specific visual features. We demonstrated how our results coincided with human fixations and also presented results that were comparable to a dominant saliency model. These results are encouraging. Consequently, we believe that with refinement our measure may be used as the foundation of a focus of attention mechanism.

For future work, we aim to use richer representations (higher-order relationships) to capture more of the low-level structure in an image. We also seek to incorporate more probabilistic principles in our measure to make it more robust. We also aim to explore our measure with regards to video sequences and depth estimation. In an effort to reduce the dimensionality and memory usage inherent with relational histograms, we aim to incorporate kernel density estimation to estimate the Rényi entropy of local pixel neighborhoods [16]. Another future goal is to implement an integrated saliency approach, incorporating both top-down and bottom-up saliency approaches to aid in visual attention and object recognition.

5. REFERENCES

[1] S. Abe. Expected Relative Entropy between a Finite Distribution and its Empirical Distribution. *SUT Journal of Mathematics*, 32:149–156, 1996.

[2] Alexander Berengolts and Michael Lindenbaum. On the Distribution of Saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1973–1987, 2006.

[3] A. P. Ashbrook, R. B. Fisher, C. Robertson, and N. Werghi. Finding Surface Correspondance for Object Recognition and Registration Using Pairwise Geometric Histograms. In *Proceedings of the 5th European Conference on Computer Vision - Volume II*, pages 674–686, 1998.

[4] T. Avraham and M. Lindenbaum. Esaliency (Extended Saliency): Meaningful Attention Using Stochastic Image Modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:693–708, 2010.

[5] N. Bruce and J. Tsotsos. Saliency Based on Information Maximization. In *Advances in Neural Information Processing Systems 18*, pages 155–162, 2006.

[6] N. D. B. Bruce and J. K. Tsotsos. Saliency, Attention, and Visual Search: An Information Theoretic Approach. *Journal of Vision*, 9:1–24, 2009.

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2009 (VOC2009). http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html.

[8] D. Gao, S. Han, and N. Vasconcelos. Discriminant Saliency, The Detection of Suspicious Coincidences, and Applications to Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:989–1005, 2009.

[9] D. Gao and N. Vasconcelos. Bottom-up Saliency is a Discriminant Process. In *IEEE 11th International Conference on Computer Vision*, pages 1–6, October 2007.

[10] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-Aware Saliency Detection. In *2010 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2376–2383, 2010.

[11] V. Gopalakrishnan, Y. Hu, and D. Rajan. Salient Region Detection by Modeling Distributions of Color and Orientation. *IEEE Transactions on Multimedia*, 11:892–905, 2009.

[12] J. S. Hare and P. H. Lewis. Scale Saliency: Applications in Visual Matching, Tracking and View-Based Object Recognition. In *Distributed Multimedia Systems 2003 / Visual Information Systems 2003*, pages 436–440, 2003.

[13] C. G. Healey, K. S. Booth, and J. T. Enns. High-speed Visual Estimation Using Peattentive Processing. *ACM*

Transactions Computer-Human Interaction, 3:107–135, 1996.
[14] X. Hou and L. Zhang. Saliency detection: A Spectral Residual Approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
[15] B. Huet and E. R. Hancock. Relational Histograms for Shape Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:1363–1370, 1999.
[16] R. Jenssen. Kernel Entropy Component Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:847 –860, 2010.
[17] T. Kadir and M. Brady. Saliency, Scale, and Image Description. *International Journal of Computer Vision*, 45:83–105, 2001.
[18] T. Kadir and M. Brady. Scale Saliency: A Novel Approach to Salient Feature and Scale Selection. In *International Conference on Visual Information Engineering*, pages 25–28, November 2001.
[19] Laurent Itti. iLab Bottom-Up Visual Attention Home Page - Example Images. http://ilab.usc.edu/bu/images/index.html.
[20] Laurent Itti. Saliency Toolbox 2.2. http://www.saliencytoolbox.net.
[21] V. Navalpakkam and L. Itti. An Integrated Model of Top-Down and Bottom-Up Attention for Optimizing Detection Speed. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2049–2056, 2006.
[22] U. Neisser. *Cognitive Psychology*. Springer-Verlag, New York, 1 edition, 1967.
[23] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape Distributions. *ACM Transactions on Graphics*, 21:807–832, 2002.
[24] A. Rényi. On Measures of Entropy and Information. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.
[25] A. Sha'ashua and S. Ullman. Structural Saliency: The Detection of Globally Salient Structures Using a Locally Connected Network. In *International Conference on Computer Vision*, pages 321–327, December 1988.
[26] Tony Lindeberg. Scale-space Theory: A Basic Tool for Analyzing Structures at Different Scales. *Journal of Applied Statistics*, 28:225–270, 1994.
[27] I. R. Vega and S. Sarkar. Statistical Motion Model Based on the Change of Feature Relationships: Human Gait-Based Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:83–105, 2003.
[28] D. Walther and C. Koch. Modeling Attention to Salient Proto-Objects. *Neural Networks*, 19:1395–1407, 2006.