TIGER: A TEXTURE-ILLUMINATION GUIDED ENERGY RESPONSE MODEL FOR ILLUMINATION ROBUST LOCAL SALIENCY

B. Chwyl, A. Chung, F. Li, A. Wong, and D. A. Clausi

Vision & Image Processing Lab, Systems Design Engineering Dept., University of Waterloo
{bchwyl, agchung, f27li, a28wong, dclausi}@uwaterloo.ca

ABSTRACT

Local saliency models are a cornerstone in image processing and computer vision, used in a wide variety of applications ranging from keypoint detection and feature extraction, to image matching and image representation. However, current models exhibit difficulties in achieving consistent results under varying, non-ideal illumination conditions. In this paper, a novel **t**exture-**i**llumination **g**uided **e**nergy **r**esponse (TIGER) model for illumination robust local saliency is proposed. In the TIGER model, local saliency is quantified by a modified Hessian energy response guided by a weighted aggregate of texture and illumination aspects from the image. A stochastic Bayesian disassociation approach via Monte Carlo sampling is employed to decompose the image into its texture and illumination aspects for the saliency computation. Experimental results demonstrate that higher correlation between local saliency maps constructed from the same scene under different illumination conditions can be achieved using the TIGER model when compared to common local saliency approach, i.e., Laplacian of Gaussian, Difference of Gaussians, and Hessian saliency models.

Index Terms— Local saliency model, illumination robust, local features, Bayesian estimation, stochastic, texture, illumination

1. INTRODUCTION

Local saliency detection in an image is a prominent research topic and a cornerstone in the field of computer vision and image processing due to its extensive use in many vision tasks. Here, local saliency refers to small, localized areas of interest that are distinct from its surrounding pixels. In contrast, global saliency (also commonly referred to as visual saliency) are larger areas of interest in the image that represent distinctive objects in a scene.

Local saliency is often used to obtain more concise representations of an image. By discarding the majority of less representative image information, the computational requirements are lowered and the robustness is improved due to the use of simple, redundant local information rather than limited complex global information [1]. As such, local saliency has become a fundamental part of many image processing and computer vision applications such as object or face recognition [2, 3], image matching [4], content based image retrieval [5], and feature keypoint detection [6, 7].

Fig. 1: Example of local saliency maps via the proposed TIGER model. Local salient regions and points (shown using a red heatmap overlay) are consistently identified for a constant scene under different illumination conditions.

In particular, local saliency is especially vital to feature detectors, which produce and match descriptors created based on locally salient keypoints in subsequent frames [8, 9, 10, 11, 6, 7, 12, 13]. Common local saliency models used for local feature keypoint detection include Laplacian of Gaussian (LoG) response, Difference of Gaussian (DoG) response, and Hessian-based response models. In LoG and DoG response models, which have been used in widely-used feature detection methods such as SIFT [6], the LoG or DoG operators (DoG is often used to approximate LoG [14], as in SIFT [6]) are applied to the image, and the LoG and DoG response map is then used to quantify the local saliency at a point in the image. In Hessian-based response models [8, 9, 11, 13], which has been used in widely-used feature detection methods such as SURF [7] and the Harris-Laplace method [11], the Hessian matrix is computed on the image, and scalar response metrics such as the determinant and the Laplacian of the Hessian matrix are used to quantify the local saliency at a point in the image. A very detailed experimental study performed in [10] found that the Hessian Laplacian model resulted in the most reliable selection of correct locally salient keypoints.

One important challenge in local saliency detection for the purpose of local feature keypoint detection is in dealing with varying illumination conditions, particularly ones that are spatially-varying (see Fig. 1 for an example of the

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Research and Innovation, and the Canada Research Chairs Program.

same scene under different illumination conditions). In practical scenarios, varying illumination conditions can affect the stability of local saliency computations, thus resulting in inconsistent keypoint detection and leading to low keypoint repeatability. Hence, a local saliency model that is robust to varying illumination conditions is highly desirable for image processing and computer vision tasks such as local feature keypoint detection.

In general, the problem of consistent local saliency detection under varying illumination is not well explored. Some methods have incorporated simple illumination compensation techniques such as intensity scaling [15] and contrast stretching [16]. Mindru *et al.* [17] aim to use colour moment invariants to account for illumination change on planar surfaces. A drawback to these methods is that their performance greatly depends on the scene. Other proposed methods use colour information to obtain better results [18, 19, 20]. Weijer *et al.* proposed a photometric invariant feature detector by combining the colour tensor with a photometric invariant derivative based on specular and diffuse models.

Motivated to address the problem of consistent local saliency detection under varying illumination at the model level, we propose a novel **t**exture-**i**llumination **g**uided **e**nergy **r**esponse (TIGER) model for illumination robust local saliency. Given an image of a scene, local saliency is quantified in the TIGER model via a modified Hessian Laplacian energy response function that is guided by a weighted aggregate of texture and illumination aspects from the image. The location of extracted locally salient areas of interest were shown to be robust to extreme illumination variations, indicating potential towards consistent local saliency if the TIGER model is used for keypoint detection.

2. METHODS

The following section outlines the methodology of the proposed TIGER model for the purpose of illumination robust local saliency to facilitate for tasks such as locally salient keypoint detection. The TIGER model aims to achieve consistent local saliency detection across various illumination conditions of the same scene by incorporating texture and illumination aspects of an image in a weighted fashion when computing the Hessian matrix to help mitigate the effects of illumination changes on the resulting local saliency response. The texture and illumination aspects of the image are approximately decomposed using a stochastic Bayesian disassociation strategy. A local saliency map is then formed based on the Laplacian of this modified Hessian.

2.1. TIGER Local Saliency Model

In Hessian-based approaches to local saliency quantification for images, the Hessian matrix Φ, is obtained for each pixel, $\bar{q}$, of an image I as follows:

$$\Phi(\bar{q}) = \begin{bmatrix} \big(\Delta_x I(\bar{q})\big)^2 & \Delta_x I(\bar{q})\Delta_y I(\bar{q}) \\ \Delta_y I(\bar{q})\Delta_x I(\bar{q}) & \big(\Delta_y I(\bar{q})\big)^2 \end{bmatrix} \tag{1}$$

where Δ_x and Δ_y are gradients in the x and y directions, respectively. The computed Hessian matrix Φ is then used to formulate a metric for local saliency (e.g., Hessian determinant, Hessian Laplacian, etc.).

A big challenge to obtaining consistent locally salient areas of interest under varying illumination conditions using this approach is that Φ is very sensitive to illumination conditions, particularly spatially-varying illumination conditions, leading to inconsistent local saliency results. Therefore, in order to produce illumination robust local saliency computations, we modify the Hessian matrix computation in the TIGER model by guiding it with a weighted aggregate of texture and illumination aspects from the image. This idea is driven by the fact that not only does the texture aspect of an image provide highly valuable information for quantifying the distinctiveness of a local area of interest relative to surrounding information, it also is less sensitive to the influence of spatially-varying illumination changes. However, despite being more sensitive to lighting variations, there is also some valuable information in the way a scene is illuminated such that using it as a weaker indicator for local saliency compared to the texture aspect is still meaningful. As such, one can control the level of guidance between texture and illumination aspects to find a balance between local saliency quality and illumination robustness.

Let us model the image I as being an additive composition of texture T and illumination L aspects:

$$I = T + L. \tag{2}$$

Given T and L, the modified Hessian matrix Φ_τ guided by a weighted aggregate of these two aspects can be defined in Eq. 3, where α and β are weights for T and L, respectively. From this modified Hessian matrix Φ_τ, the local saliency, s, at $\bar{q}$ can be computed as a modified Hessian energy response defined by the determinant and trace of the Laplacian of the Hessian matrix [21]:

$$s(\bar{q}) = \frac{det\big(\Phi_\tau(\bar{q})\big)}{trace\big(\Phi_\tau(\bar{q})\big)}. \tag{4}$$

By repeating this process for each pixel within $\bar{I}$, a local saliency map can be produced.

2.2. Bayesian Disassociation

To be able to compute the modified Hessian matrix Φ_τ, we must determine the texture (T) and illumination (L) aspects of an image. Based on the model in Eq. 2 that relates T, L, and I in an additive manner, we aim to disassociate T and L using a Bayesian dissociation approach, where we first compute an

$$\Phi_\tau(\bar{q}) = \begin{bmatrix} \left(\Delta_x\{\alpha T + \beta L\}(\bar{q})\right)^2 & \Delta_x\{\alpha T + \beta L\}(\bar{q})\Delta_y\{\alpha T + \beta L\}(\bar{q}) \\ \Delta_y\{\alpha T + \beta L\}(\bar{q})\Delta_x\{\alpha T + \beta L\}(\bar{q}) & \left(\Delta_y\{\alpha T + \beta L\}(\bar{q})\right)^2 \end{bmatrix} \quad (3)$$

approximation of L (denoted here as $\hat{L}$) and then approximate T as the residual between I and $\hat{L}$ (i.e., $\hat{T} = I - \hat{L}$). Let us formulate the problem of obtaining $\hat{L}$ given I as a Bayesian least-squares minimization problem based on Eq. 2:

$$\begin{aligned}\hat{L} &= \arg\min_{\hat{L}} E((L-\hat{L})^2|I) \\ &= \arg\min\nolimits_{\hat{L}} \left(\int (L-\hat{L})^2 p(L|I)d(L)\right), \end{aligned} \quad (5)$$

where $E(.)$ denotes the expectation. Taking the derivative of Eq. 5, setting it equal to zero, and expanding yields the following equation:

$$\int Lp(L|I)d(L) = \int \hat{L}p(L|I)d(L). \quad (6)$$

since

$$\hat{L} = \int \hat{L}p(L|I)d(L), \quad (7)$$

substituting Eq. 7 into Eq. 6 gives us the final expression:

$$\hat{L} = \int Lp(L|I)d(L). \quad (8)$$

Note that the posterior probability, $p(L|I)$, is required to compute Eq. 8. However, it is difficult to obtain $p(L|I)$ in an analytical manner. For this reason, a non-parametric approach is used to obtain $p(L|I)$ via a Monte Carlo sampling strategy.

2.3. Posterior Probability Estimation

To estimate $p(L|I)$, a Monte Carlo sampling approach was applied. We first aim to establish a set of pixels, Ω, within a region of interest, $\eta_{\bar{q}}$, surrounding the pixel of interest, $\bar{q}$. A uniform distribution, $Q(q_k, \bar{q})$, is used as an instrumental distribution to sample pixels with equal probability. Upon selecting a pixel, an acceptance probability, $\alpha(q_k|\bar{q})$, is calculated based on regional similarity between the sample pixel, q_k, and the pixel of interest, $\bar{q}$, as follows:

$$\alpha(q_k|\bar{q}) = exp\left(\sigma - \frac{1}{N}\sum_{i=1}^{N}\left(\aleph_{q_k}(i) - \aleph_{\bar{q}}(i)\right)^2\right) \quad (9)$$

where $\aleph_{q_i}$ and $\aleph_{\bar{q}}$ represent regions of equal size surrounding sample pixel q_k and $\bar{q}$ respectively, and N represents the number of pixels within each region. The acceptance probability determines the likelihood that q_k is added to the set Ω. The process of sampling pixels from $\eta_{\bar{q}}$ is repeated until M sample pixels are acquired. The posterior probability can then be calculated as

$$\hat{p}(L|I) = \frac{\sum_{k=0}^{M}\alpha(q_k|\bar{q})\delta\big(L - I(q_k)\big)}{Z}, \quad (10)$$

where δ is the Dirac function and Z is used as a normalization factor to ensure $\sum \hat{p}(L|I) = 1$.

3. RESULTS

3.1. Experimental Setup

To test the efficacy of the proposed TIGER model for obtaining consistent local saliency results against drastic illumination conditions, images of a constant scene captured in a windowless room using a Sony HDR-AS30V camera under different lighting angles were used. A scene of assorted objects was assembled in front of a solid background, and illumination was provided by a spotlight. To alter illumination, the spotlight was moved about the room as images were captured. In addition, selections from the Yale Face Database B [22], the GTILT dataset [23], and the AMOS dataset [24] were used. Each image set consisted of a single scene under varying illumination, and a total of 49 images across 10 different scenes was used [1].

To better assess the performance of TIGER with respect to commonly-used local saliency models employed by local feature detection methods, TIGER was compared to the Laplacian of Gaussian (LoG) [14], Difference of Gaussians (DoG) [6], and the Hessian [9] approaches using the image sets. All methods were evaluated using a correlation metric calculated between local saliency maps of the same scene under different lighting conditions. A high correlation score is desirable, and implies a high level of consistency among the saliency maps despite varying illumination. Based on empirical testing, the α and β values for TIGER were determined to be 0.6 and 0.4, respectively.

3.2. Experimental Results

For each image set, the proposed TIGER model and other tested approaches were used to compute local saliency maps of the same scene under different lighting conditions. The methods were assessed visually for consistent salient region identification across varying illumination. Figure 2 shows the identified salient regions for each method for a single scene.

Looking at Figure 2, the local saliency maps produced using LoG [14] and DoG [6] show a tendency towards identifying edges as salient rather than compact regions or corners. While useful for applications such as segmentation, edges are often insufficiently unique to be used as local features, e.g.,

[1]Available: http://vip.uwaterloo.ca/people/brendan-chwyl/VICS.html

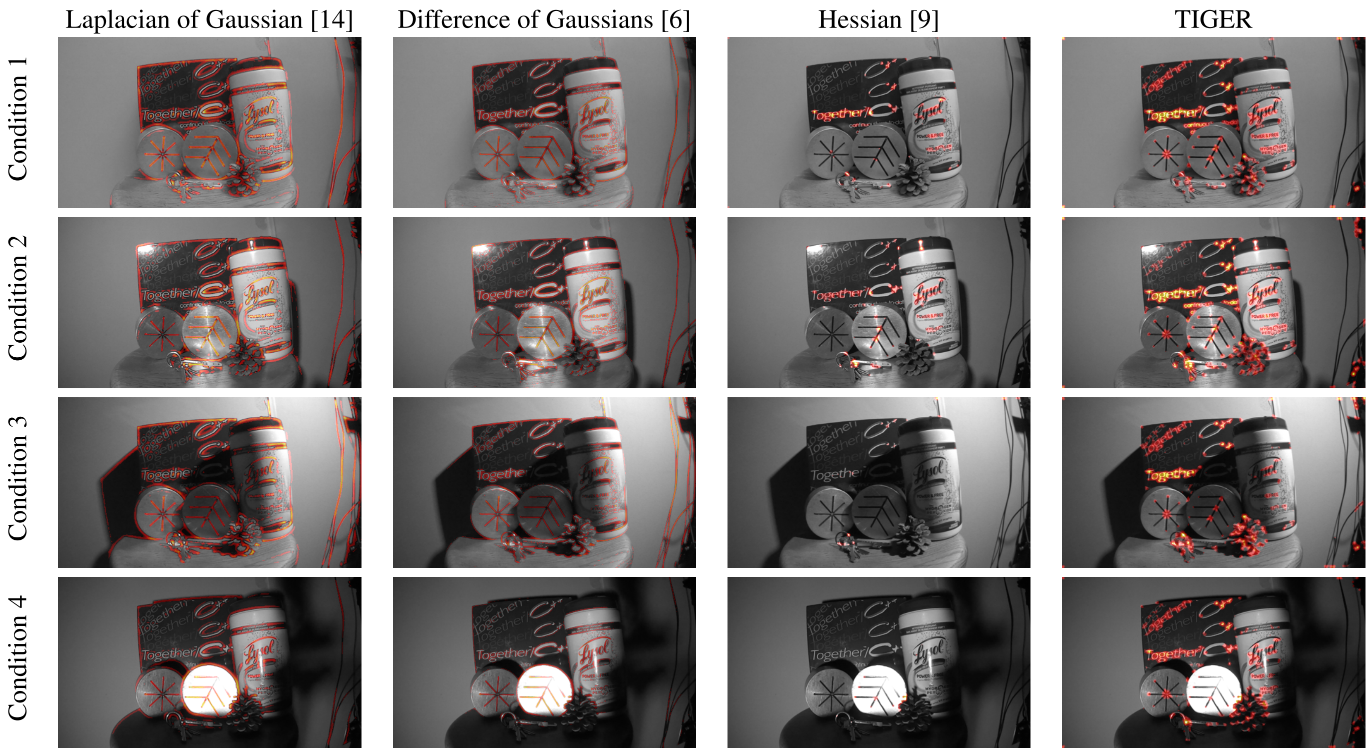

Fig. 2: Comparison of the Laplacian of Gaussian, Difference of Gaussians, Hessian, and TIGER models of the same scene under four different illumination conditions. The scenes are shown with their corresponding saliency maps overlayed in red heatmap, where a lighter colour indicates a higher local saliency.

keypoints, as such require additional processing to obtain keypoints [6]. It can also be observed that different illumination conditions greatly affect the local saliency of identified edges. For example, it can be seen that the results produced by DoG and LoG show noticeably weaker edges for Condition 4.

The Hessian [9] and TIGER models generally identify high-contrast corner points or compact regions as locally salient, as can be seen in Figure 2 where the Hessian and TIGER both highlight the text near the centre of the image. While the Hessian approach clearly identifies local salient areas of interest in well-lit images, the approach begins to show its limitations in more drastic illumination conditions (such as Condition 3 or 4 in Figure 2). Figure 2 also shows that the local saliency maps produced by TIGER has the most visually consistent set of locally salient areas of interest across the different lighting conditions.

In addition to visual assessment, the tested local saliency models were evaluated for each image set using a correlation metric; that is, the correlation between local saliency maps for a constant scene was calculated to quantitatively measure the consistency of identified salient regions. The correlation metric was then averaged across all the image sets to produce an overall measure, as shown in Table 1. It can be observed that while the overall correlation of each approach is relatively low, TIGER has the most consistent local saliency maps (with 53.4%) given drastic illumination conditions.

Table 1: Overall saliency map correlation for LoG, DoG, Hessian, and TIGER. TIGER achieved the highest overall correlation, indicating the highest level of consistency across identified locally salient regions for a constant scene.

Local Saliency Model	Overall Correlation
LoG [14]	43.4%
DoG [6]	38.5%
Hessian [9]	33.5%
TIGER	**53.4%**

4. CONCLUSION

In this paper, we presented a novel **t**exture-**i**llumination **g**uided **e**nergy **r**esponse (TIGER) model for an illumination robust local saliency. TIGER was compared to other commonly local saliency models using images with varying lighting angles of 10 different scenes. A correlation metric and visual assessment of the experimental results indicate that TIGER provides more consistent local saliency results given drastic lighting illumination conditions, and is a good candidate for use for illumination robust feature detection. Future works include more comprehensive testing and extending TIGER for use in global saliency scenarios [25].

5. REFERENCES

[1] Bill Triggs, "Detecting keypoints with stable position, orientation, and scale under illumination changes," in *Computer Vision-ECCV 2004*, pp. 100–113. Springer, 2004.

[2] Krystian Mikolajczyk, Bastian Leibe, and Bernt Schiele, "Multiple object class detection with a generative model," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE, 2006, vol. 1, pp. 26–36.

[3] Ajmal S Mian, Mohammed Bennamoun, and Robyn Owens, "Keypoint detection and local feature matching for textured 3d face recognition," *International Journal of Computer Vision*, vol. 79, no. 1, pp. 1–12, 2008.

[4] Avi Kelman, Michal Sofka, and Charles V Stewart, "Keypoint descriptors for matching across multiple image modalities and non-linear intensity variations," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–7.

[5] Pedro Martins, Paulo D Carvalho, and Carlo Gatta, "Context aware keypoint extraction for robust image representation.," in *BMVC*, 2012, pp. 1–12.

[6] David G Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Ieee, 1999, vol. 2, pp. 1150–1157.

[7] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[8] Stephen M Smith and J Michael Brady, "Susana new approach to low level image processing," *International journal of computer vision*, vol. 23, no. 1, pp. 45–78, 1997.

[9] Chris Harris and Mike Stephens, "A combined corner and edge detector.," in *Alvey vision conference*. Manchester, UK, 1988, vol. 15, p. 50.

[10] Krystian Mikolajczyk and Cordelia Schmid, "Indexing based on scale invariant interest points," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. IEEE, 2001, vol. 1, pp. 525–531.

[11] Krystian Mikolajczyk and Cordelia Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[12] Stefan Leutenegger, Margarita Chli, and Roland Yves Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2548–2555.

[13] Edward Rosten, Reid Porter, and Tom Drummond, "Faster and better: A machine learning approach to corner detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 1, pp. 105–119, 2010.

[14] Tony Lindeberg, "Feature detection with automatic scale selection," *International journal of computer vision*, vol. 30, no. 2, pp. 79–116, 1998.

[15] Bogdan Georgescu and Peter Meer, "Point matching under large image deformations and illumination changes," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 6, pp. 674–688, 2004.

[16] Murat Gevrekci and Bahadir K Gunturk, "Illumination robust interest point detection," *Computer Vision and Image Understanding*, vol. 113, no. 4, pp. 565–571, 2009.

[17] Florica Mindru, Theo Moons, and Luc Van Gool, "Comparing intensity transformations and their invariants in the context of color pattern recognition," in *Computer VisionECCV 2002*, pp. 448–460. Springer, 2002.

[18] Joost Van De Weijer, Theo Gevers, and Arnold WM Smeulders, "Robust photometric invariant features from the color tensor," *Image Processing, IEEE Transactions on*, vol. 15, no. 1, pp. 118–127, 2006.

[19] Ranjith Unnikrishnan and Martial Hebert, "Extracting scale and illuminant invariant regions through color," *Robotics Institute*, p. 356, 2006.

[20] Flore Faille, "Stable interest point detection under illumination changes using colour invariants.," in *BMVC*, 2005.

[21] J Alison Noble, "Finding corners," *Image and Vision Computing*, vol. 6, no. 2, pp. 121–128, 1988.

[22] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[23] M. R. Bales, D. Forsthoefel, D. S. Wills, and L. M. Wills, "The Georgia Tech Illumination Transition (GTILT) Dataset," Mobile Vision Embedded Systems Lab (MoVES), Georgia Institute of Technology, http://www.ece.gatech.edu/research/labs/pica/GTILT.html, November 2012.

[24] Nathan Jacobs, Nathaniel Roman, and Robert Pless, "Consistent Temporal Variations in Many Outdoor Scenes," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun 2007.

[25] Christian Scharfenberger, Alexander Wong, Khalil Fergani, John S Zelek, and David A Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 979–986.