ioM-HDR: A High Dynamic Range adapted model of visual salience

Andre Harrison[1[0000-0002-6850-9677]*], Michael Green[1], Chou Hung[2[0000-0003-1447-9828]], Adrienne Raglin[1]

[1] CCDC Army Research Laboratory, Adelphi, MD 20783, USA
[2] CCDC Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA
{andre.v.harrison2.civ, michael.a.green85.ctr, chou.p.hung.civ, adrienne.j.raglin.civ}@mail.mil

Abstract. Visual salience is the visual percept from an image or scene that attracts a person's attention and models of visual salience are often used to study visual search. Visual salience is often used to model human gaze patterns, but it can also be utilized to prioritize the processing of a scene by autonomous systems in a way that is consistent with human behavior. However, current models of visual search behavior are based on laboratory tests where luminance contrast ratios don’t exceed 1000:1 (‘low dynamic range’, LDR), but visual search in the real world occurs under luminance contrast ratios of up to 1,000,000:1. In this paper we present a high dynamic range adapted model of our ideal observer model of visual salience as a way of overcoming the dynamic range limitations of other low dynamic range models of visual salience.

Keywords: visual salience, tone mapping, high dynamic range, entropy

1 Introduction

At a given moment in time the maximum relative difference in luminance that a person can perceive, a person’s in-scene dynamic range, is on the order of 10^5:1. Human vision has such a high dynamic range because the outside world is a high dynamic range (HDR) one, where changes in luminance (contrast) are often greater than 10^3:1. However, computer vision algorithms continue to be designed and evaluated assuming the dynamic range of the environment is low, where the maximum dynamic range, in luminance, of the environment is assumed to be 10^3:1 or less. Thus, computer vision algorithms are almost exclusively tested on imagery that only uses 8-bits/color channel/pixel. This bit-depth is appropriate for indoor environments and even many outdoor environments when there is no large change in luminance (no deep shadows or regions of bright glare). But as autonomous military systems operate more and more outside of controlled laboratory settings and in more real-world environments with higher dynamic ranges this discrepancy in representation is likely to have a negative impact on their reliability and mission performance. Some recent papers have looked into the performance of basic elements of computer vision algorithms (Canny edge detectors, blob detectors, SIFT/SURF feature detectors) and many have often found a consistent degradation in the performance of these algorithms on HDR imagery if no form of adaption is used [1–4].

Within the space of computer vision algorithms, models of visual salience are somewhat unique in how closely tied they are to human perception. Visual salience is an aspect of perception that makes some objects or locations within a scene stand out or demand more attentional resources compared to others [5, 6]. In the human visual system, it is used to prioritize the limited gaze and attentional resources that a person can use to perceive their environment. For autonomous systems, computational models of visual salience can be used as a way to implement human consistent methods of visual search by predicting how likely it is that each location within an image will attract a person's gaze. This also allows autonomous agents to prioritize visual information in a way that is understandable by humans, which may improve feelings of trust in the systems and overall teamwork.

However, high dynamic range stimuli can negatively affect the performance of visual saliency models for two major reasons. Like most computer vision algorithms, models of visual salience have been developed and evaluated on LDR, 24-bit color, stimuli, but these models have only recently been evaluated on high dynamic range stimuli to see how well they can maintain predictive accuracy [7]. The other issue that computational models of visual salience face that is somewhat unique within computer vision is that these models are founded on studies of human perception, but these studies were conducted on display systems that could only support a limited range of luminance levels. As such, theories of visual perception and attention have been based on behaviors on LDR stimuli and were just assumed to continue to operate in HDR environments. Only in the last few years have displays or projectors become commercially available with dynamic ranges high enough to show uncompressed HDR imagery to test those assumptions. And recent studies using these HDR capable displays or projectors have begun to overturn some of the assumptions founded on LDR systems, though the exact extent of the implications of these findings still needs to be fully explored.

There have been several approaches employed by other research to adapt existing computer vision algorithms to handle HDR image data by manipulating the imagery to appear more similar to an LDR image [8, 9]. Other approaches that have been taken primarily in the visual salience research area try to implement adaptation methods found within the HVS and have incorporated those into a visual saliency model with moderate results [10, 11]. In this paper, we present our approach to modeling visual salience in a way that can adapt to images with any range of luminance levels by taking inspiration from work done in the area of tone mapping.

2 Ideal Observer Model: High Dynamic Range

2.1 Adaptation to HDR environments

Within computer graphics and models of human vision, the brightness of an object is typically modeled as the combination of the reflectance of light off of that object and the illumination level in that location from a set of light sources. And it is the variation in illumination within a scene that largely determines the dynamic range of that scene, while most of the information used by models of vision are based on an object’s reflectance information. Some tone mapping methods use this fact to separate an image into an illumination layer and a reflectance layer so that the illumination layer can be compressed without affecting the reflectance layer. Tone mapping methods are HDR compression algorithms that aim to compress the dynamic range of an HDR image so that it can be displayed on an LDR display (i.e. an HDR image is compressed to be an LDR 24-bit color image), while trying to preserve some aspect of the original HDR image. In order to make our visual saliency model adaptable to images with larger dynamic ranges, we use this light model layer separation approach in the form of applying a bilateral filter to the input image to try and separate the reflectance and illumination layers. The fundamental equation of the bilateral filter is that of an edge preserving filter used to estimate the illumination layer of an image as shown in eqn. (1). It allows the creation of a low pass version of the image without blurring the lines of object boundaries. The filter works by computing a weighted average of a group of pixels *Ω* surrounding pixel *s*, where the weighted contribution of each pixel *pϵΩ* is proportional to the distance of between pixel *p* and *s* as well as the difference in value ($I_p - I_s$). Thus, pixels that are further away from pixel *s* or have a large difference in luminance value contribute a small weight to the weighted average. To keep the computational speed acceptable, we use the implementation of the bilateral filter by Ghosh and Chaudhury that has an O(1) computational cost [12, 13].

$$\begin{gathered} J_s = \frac{1}{k(s)} \sum_{p \in \Omega} f(p-s)\, g\left(I_p - I_s\right) I_p \\ \textit{and}\ \mathrm{k}(s) = \sum_{p \in \Omega} f(p-s)\, g\left(I_p - I_s\right) \end{gathered} \tag{1}$$

2.2 Updated ioM

The ideal observer model (ioM) as it was originally proposed in [14] applied a multi-resolution spatially-oriented wavelet decomposition to each feature channel of an image (color and intensity). It then tried to calculate the entropy, H(X), of each region within each subband of the wavelet decomposition by modeling each coefficient as a Gaussian Scale Mixture. However, this formulation was analytically unsolvable and so required several numerical approximations in order to implement it. In this current formulation, rather than assuming the wavelet coefficients are part of a zero mean Gaussian Scale Mixture, we model them using a zero mean Generalized Gaussian Distribution eqn. (2), with shape and scale parameters (*r, s*), respectively.

$$P_{s,r}(x) = \frac{r}{2s\Gamma\left(\frac{1}{r}\right)} e^{-\left|\frac{x}{s}\right|^{r}} \tag{2}$$

The Generalized Gaussian Distribution (GGD) is a family of probability distribution functions that depending on the value of the shape parameter can model different probability distributions. For r=1, (2) takes the form of the Laplacian distribution, while if r=2, (2) is a zero mean normal distribution with a variance given by s^2. *Γ(·)* from eqn (2) is the gamma function. For natural images the distribution of wavelet coefficients is typically peaky and highly kurtotic, which can be modeled using a shape parameter with an r<1. For this range of shape parameters, when the shape parameter needs to be estimated using a few samples, ***x***, the negative entropy method (J) for estimating the shape parameter can produce accurate results [15, 16]. The equation for the negative entropy method is given in eqn. (3), but inverting it to solve for r is not easy and may not even be possible, but r can be found through the use of a look up table and subsequent interpolation to improve the precision depending on precision requirements. With r solved, the shape parameter s can be estimated using eqn. (4) [17].

$$\Psi(r) = \log\left(\frac{r}{2}\sqrt{\frac{\Gamma(0.5)^{3}\,\Gamma\left(\frac{3}{r}\right)}{\Gamma\left(\frac{1}{r}\right)^{3}\,\Gamma(1.5)}}\right) + \left(0.5-\frac{1}{r}\right) = J$$

$$\textit{and}\ \ J(\mathbf{x}) = k_1\left[E\left\{\mathbf{x}e^{\left(\frac{-\mathbf{x}^2}{2}\right)}\right\}\right]^2 + k_2\left[E\left\{e^{\left(\frac{-\mathbf{x}^2}{2}\right)}\right\} - \sqrt{\frac{1}{2}}\right]^2 \tag{3}$$

Where k_1 = 7.412 and k_2 = 33.67 and r= Ψ^{-1}(J)

$$\hat{s} := \left(\frac{\hat{r}}{n}\sum_{i=1}^{n}\left|x_i\right|^{\hat{r}}\right)^{\frac{1}{\hat{r}}} \tag{4}$$

Where $\sum_{i=1}^{n} x_i = \mathbf{x}$ is the set of samples, n, used to estimate the GGD parameters.

Incorporating the new underlying mathematical model for the ideal observer model and the bilateral filter to adapt to images with high luminance ranges, we show the diagram of the ioM-HDR in Figure 1. The feature channel separation step of the ioM-HDR model decomposes an HDR image into an illumination map, a reflectance map, and a normalized color map. The illumination map is the output of the bilateral filter after it has been applied to the luminance version of the HDR image. The color and reflectance maps are then created by normalizing the color and luminance maps of the HDR image by the illumination layer. During the wavelet decomposition and entropy calculation, each map goes through a multi-resolution wavelet decomposition using oriented spatial filters to create subband maps at each resolution level for each feature.

For each region within a subband map, the entropy of that region is estimated using the GGD to model the probability distribution of the samples in that region. Once entropy maps for each subband are created, the maps are normalized, using the normalized method from the original Itti model [18], to look for unique patterns within each entropy map. The normalized entropy maps are then averaged together across subbands within a feature channel and then across all feature channels. The final averaged normalized output is the resulting saliency map whose values serve as a likelihood estimate of how likely each location will attract a person's gaze.

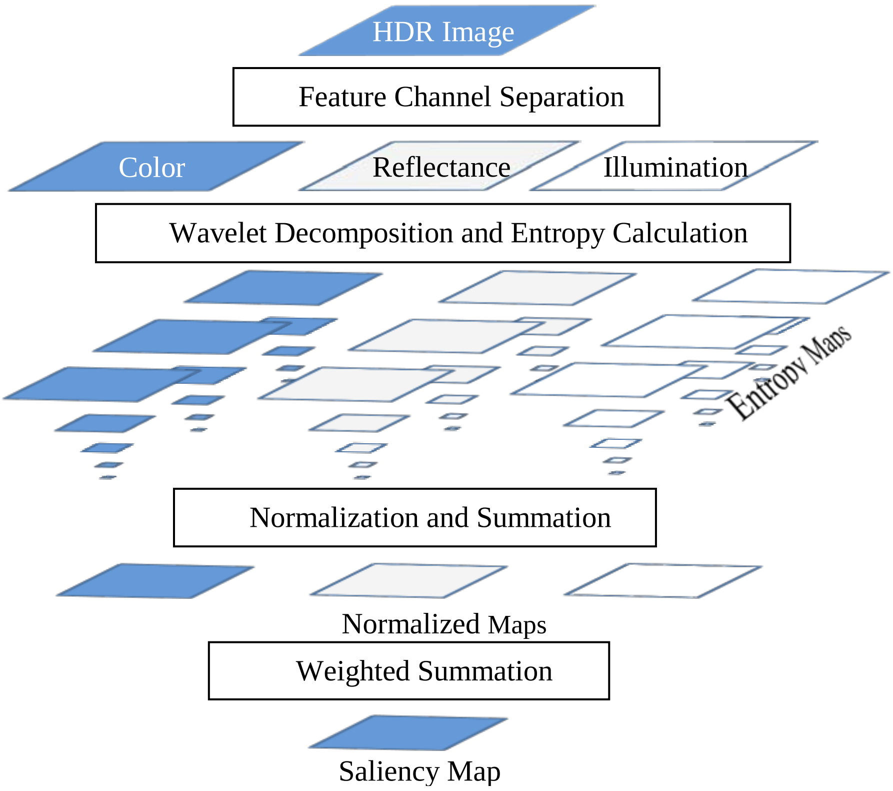
HDR Image
Feature Channel Separation
Color
Reflectance
Illumination
Wavelet Decomposition and Entropy Calculation
Entropy Maps
Normalization and Summation
Normalized Maps
Weighted Summation
Saliency Map

Fig. 1. Diagram of the High Dynamic Range extension of the Ideal Observer Model.

3 Conclusions

As autonomous military systems and platforms move from tools to teammates to fully autonomous platforms and systems they will have to become more adapted to operating in new and dynamically changing environments in order to efficiently achieve mission objectives while also maintaining an awareness of their surrounding environment. The ability of autonomous platforms to process HDR stimuli will not just support both of these tasks, but will be a fundamental requirement of these systems as every moment information within the environment remains undetected, and hence unprioritized, adds danger to the mission and risks mission success. The model presented in this paper takes a step towards that aim, by designing a model of visual salience that is able to predict eye gaze while automatically adapting to the dynamic range of the input image. Further testing and development of the model will allow a more quantitative demonstration of the predictive performance of this model.

REFERENCES

1. Rana, A., Valenzise, G., Dufaux, F.: Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions. In: 2015 IEEE International Symposium on Multimedia (ISM). pp. 289–294. IEEE (2015). https://doi.org/10.1109/ISM.2015.58.
2. Rana, A., Valenzise, G., Dufaux, F.: An evaluation of HDR image matching under extreme illumination changes. In: 2016 Visual Communications and Image Processing (VCIP). pp. 1–4. IEEE (2016). https://doi.org/10.1109/VCIP.2016.7805556.
3. Přibyl, B., Chalmers, A., Zemčík, P., Hooberman, L., Čadík, M.: Evaluation of feature point detection in high dynamic range imagery. J. Vis. Commun. Image Represent. 38, 141–160 (2016). https://doi.org/10.1016/j.jvcir.2016.02.007.
4. Chermak, L., Aouf, N.: Enhanced feature detection and matching under extreme illumination conditions with a HDR imaging sensor. In: 2012 IEEE 11th International Conference on Cybernetic Intelligent Systems (CIS). pp. 64–69. IEEE (2012). https://doi.org/10.1109/CIS.2013.6782161.
5. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cogn. Psychol. 12, 97–136 (1980). https://doi.org/10.1016/0010-0285(80)90005-5.
6. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res. 40, 1489–1506 (2000). https://doi.org/10.1016/S0042-6989(99)00163-7.
7. Harrison, A. (U. S.A.R.L., Green, M., Hung, C., Raglin, A.J.: Predicting where people look in high dynamic range images: VS in HDR images what’s missing? In: IEEE International Workshop on Multimedia Signal Processing (2019).
8. Wang, W., Wang, Y., Huang, Q., Gao, W.: Measuring visual saliency by Site Entropy Rate. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 2368–2375. IEEE (2010). https://doi.org/10.1109/CVPR.2010.5539927.
9. Zhang, J., Yu, M., Song, Y., Shao, H., Jiang, G.: Luminance regionalization-based saliency detection for high dynamic range image. In: Dai, Q. and Shimura, T. (eds.) Proc. SPIE 10817, Optoelectronic Imaging and Multimedia Technology V. p. 43. SPIE (2018). https://doi.org/10.1117/12.2502042.

10. Bremond, R., Petit, J., Tarel, J.-P.: Saliency Maps of High Dynamic Range Images. In: Kutulakos, K.N. (ed.) European Conf. on Comp. Vision. pp. 118–130. Springer Berlin Heidelberg (2012). https://doi.org/10.1007/978-3-642-35740-4_10.
11. Dong, Y., Pourazad, M.T., Nasiopoulos, P.: Human Visual System-Based Saliency Detection for High Dynamic Range Content. IEEE Trans. Multimed. 18, 549–562 (2016). https://doi.org/10.1109/TMM.2016.2522639.
12. Ghosh, S., Chaudhury, K.N.: On fast bilateral filtering using fourier kernels. IEEE Signal Process. Lett. 23, 570–574 (2016). https://doi.org/10.1109/LSP.2016.2539982.
13. Chaudhury, K.N.: Acceleration of the shiftable O(1) algorithm for bilateral filtering and nonlocal means. IEEE Trans. Image Process. 22, 1291–1300 (2013). https://doi.org/10.1109/TIP.2012.2222903.
14. Harrison, A., Etienne-Cummings, R.: An entropy based ideal observer model for visual saliency. In: 2012 46th Annual Conference on Information Sciences and Systems (CISS). pp. 1–6. IEEE (2012). https://doi.org/10.1109/CISS.2012.6310928.
15. Yu, S., Zhang, A., Li, H.: A Review of Estimating the Shape Parameter of Generalized Gaussian Distribution⋆. J. Comput. Inf. Syst. 8, 9055–9064 (2012).
16. Prasad, R., Saruwatari, H., Shikano, K.: Estimation of shape parameter of GGD function by negentropy matching. Neural Process. Lett. 22, 377–389 (2005). https://doi.org/10.1007/s11063-005-1385-9.
17. Song, K.S.: Globally convergent algorithms for estimating generalized gamma distributions in fast signal and image processing. IEEE Trans. Image Process. 17, 1233–1250 (2008). https://doi.org/10.1109/TIP.2008.926148.
18. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1254–1259 (1998). https://doi.org/10.1109/34.730558.