Virtual Photograph Based Saliency Analysis of High Dynamic Range Images

Xihe Gao* Stephen Brooks† Dirk V. Arnold‡
Faculty of Computer Science
Dalhousie University
Halifax, Nova Scotia, Canada

Abstract

Computational visual attention systems detect regions of interest in images. These systems have a broad range of applications in areas such as computer vision, computational aesthetics, and non-photorealistic rendering. However, almost all the systems to date are designed for low dynamic range (LDR) images and may not be suitable for analyzing saliency in high dynamic range (HDR) images. We propose a novel algorithm for saliency analysis of HDR images that is based on virtual photographs. Taking virtual photographs is the inverse process of generating HDR images from multiple LDR exposures, and the virtual photograph sequence has the capacity to more comprehensively reveal salient content in HDR images. We demonstrate that our method can produce more consistently reliable results than existing methods.

CR Categories: I.4.0 [Image Processing and Computer Vision]: General—Image processing software; I.4.7 [Image Processing and Computer Vision]: Scene Analysis;

Keywords: HDR images, saliency analysis, virtual photographs

1 Introduction

The human brain tends to prioritize visual data obtained from the real world to guide our gaze, which is known as selective attention [Frintrop et al. 2010]. Based on this mechanism, computational visual attention systems have been designed to detect regions of interest in digital images and are widely used in computer vision, computational aesthetics, and non-photorealistic rendering methods.

Although existing visual attention systems can lead to satisfying results for LDR images, they do not always perform well when applied to HDR images [Brémond et al. 2012]. Compared with LDR images, HDR images allow a much higher dynamic range to represent luminance in the real world. However, if computational visual attention systems are applied to HDR images directly, the dynamic range will be linearly scaled, which can lead to the loss of HDR content that in turn makes salient regions appear not salient or vice versa. Details and textures within HDR images may not register as salient due to the significant contrast reduction, preventing visual attention systems from obtaining useful results. At the same time, Narwaria et al. [2012] have found that tone mapping operators "can [. . .] modify human attention and fixation behavior significantly", thus rendering the approach of applying salience analysis techniques after dynamic range compression unreliable.

*e-mail:xgao@cs.dal.ca
†e-mail:sbrooks@cs.dal.ca
‡e-mail:dirk@cs.dal.ca

CAe 2013, July 19 – 21, 2013, Anaheim, California.

In this paper, we propose a novel algorithm for computing saliency maps of HDR images by taking sequences of virtual photographs with different exposure settings. To generate the virtual photographs, we use a generic camera response function mapping radiances to pixel intensities in combination with a simple yet effective approach for calibrating for varying ambient luminance. A salience detection algorithm is applied to each of the virtual photographs, and the resulting salience maps are combined to form the salience map of the HDR image. We present evidence that the areas of visual attention that are identified are more meaningful than those identified by other approaches.

2 Background

Most computational attention systems, such as those proposed by Itti et al. [1998] or Frintrop [2006], are built on feature integration theory as outlined by Treisman and Gelade [1980]. The core idea is to detect several types of features and combine their saliencies to generate a saliency map. The model of Itti et al. [1998] was revised by Itti and Koch [2000], who adopt iterative normalization for within-feature competition to replace the simple normalization used in the earlier paper.

In the work of Itti and Koch [2000], the visual attention model is designed to predict bottom-up attention, which is derived only from the visual scene in a static color image. Multiple features, including intensity, color, and orientation, are combined to produce saliency or conspicuity maps. The first step is to extract early visual features with image pyramids. Each feature is computed by center-surround mechanisms, also known as center-surround differences, which compare the value of center region and surround region in the receptive fields [Frintrop et al. 2010]. The operation is implemented as the difference between fine and coarse scales. Then, cross-scale information is combined to produce maps of features. A within-feature spatial competition scheme is used to solve the signal-to-noise problem, and the interaction is realized by a two-dimensional difference-of-Gaussians approach. Finally, the feature maps are summed into a single saliency map of the image. Besides that map, the model also computes the trajectory of visual attention as an output, in which a winner-take-all network is used to select image regions with local saliency maxima. A fixed size disk is applied to represent the focus of attention since the visual attention is usually on a region rather than a single point [Frintrop et al. 2010].

The approach of Itti and Koch [2000] can achieve satisfactory performance when applied to the problem of detecting salient regions in LDR images, and therefore is widely applied to saliency-based research, such as object detection [Viola and Jones 2004; Mohan et al. 2001; Walther and Koch 2006], image resizing [Avidan and Shamir 2007; Wang et al. 2008], non-photorealistic rendering [Dong et al. 2013; Lee et al. 2013], video compression [Itti 2004] as well as tone mapping of HDR images [Lin and Yan 2011]. The technique is regarded as one of the most influential models for computational saliency analysis [Frintrop et al. 2010] and we use it as the basis for our approach.

Derived from the visual attention system by Itti et al. [1998], Brémond et al. [2012] propose a method for computing saliency

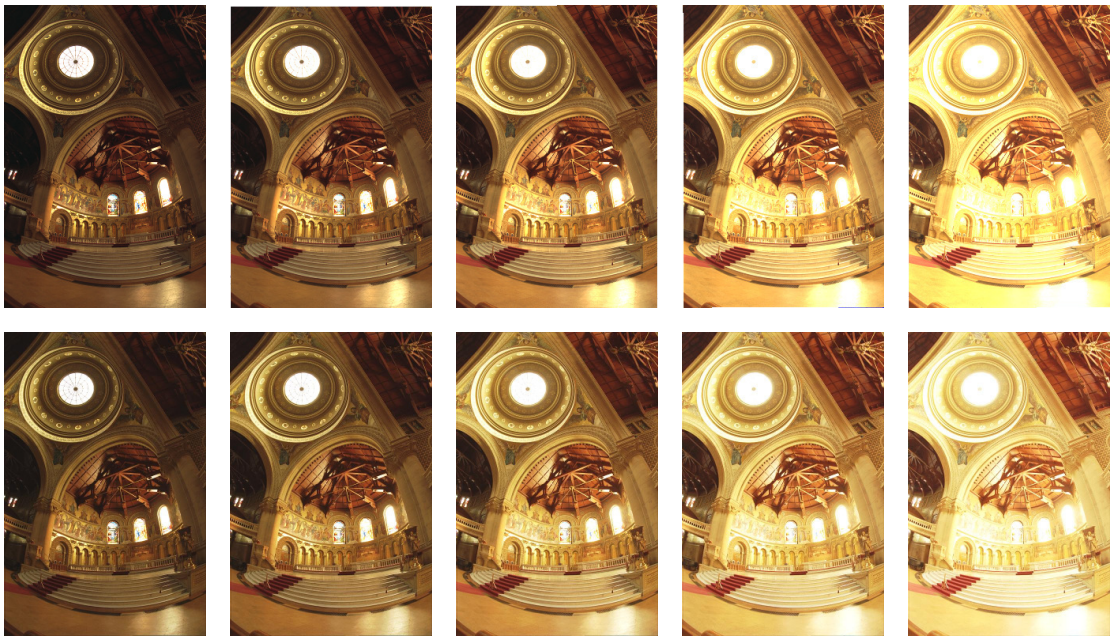

Figure 1: *Comparison between real photographs (upper row) and virtual photographs (lower row). Exposure times in all cases increase by a factor of two between neighboring images when moving from left to right. The center image in the lower row has an exposure time of $\Delta t = 1$. The HDR image that the virtual photographs are derived from as well as the real photographs are due to Paul Debevec, © 1997 ACM.*

maps of HDR images. Based on the observation that saliency is better preserved in color maps, they modify the definition of the other visual features. Specifically, intensity and orientation differences are normalized by division by the intensity. The authors suggest that the normalization of intensity "may be seen as a gain modulation, which is the physiological mechanism of visual adaptation". In eye tracking experiments with human subjects, they find that their approach provides more accurate saliency maps than that of Itti and Koch [2000] when the latter is applied either directly to the HDR image or after dynamic range compression with one of six tone mapping operators. We will use the approach of Brémond et al. [2012] as a baseline to compare our algorithm against.

3 Algorithm

Given an HDR image, our method computes a sequence of virtual photographs of that image, analyzes the saliency map of each virtual photograph, and subsequently combines them into a unified saliency map. We discuss the taking of virtual photographs in Section 3.1. In Section 3.2, we describe how to combine the saliency maps of the virtual photographs into that of the HDR image.

3.1 Taking Virtual Photographs

Taking virtual photographs was first introduced for flash-exposure HDR imaging [Agrawal et al. 2005; Richardt 2008], in which a series of images with different flash intensity and exposures are captured and merged into HDR maps and virtual photographs can then be taken for any combination of exposure and flash intensity. In these methods, the response function of the camera that is recovered when producing an HDR image is applied for taking virtual photographs from the same HDR image. Consequently, it is assumed that the response function is already available.

For HDR images that are captured directly by a digital camera, generated synthetically using ray tracing, or generated using other approaches different from those that combine multiple LDR images, a response function may not easily be available. In order to obtain well-exposed virtual photographs from an HDR image if the response function of the capturing device is not available, we use self-calibration based on ambient luminance in combination with a generic response function.

In photography, the exposure of a scene is determined with reference to ambient lighting, which can be measured by light meter strategies such as average metering or spot metering. The metered light is assumed to have 18% reflectance and is recognized as middle gray on the lightness scale [Brown 2011]. Our calibration algorithm is inspired by this process. Given an HDR image with luminance values $L(x, y)$, we compute the average logarithmic luminance as

$$L_{av} = \exp\left(\frac{1}{N}\sum_{x,y}\ln\left(L(x,y)\right)\right) \tag{1}$$

where the sum extends over all pixels and N is the total number of pixels, and refer to it as ambient luminance. The ambient luminance should be displayed as middle gray on the lightness scale. In the sRGB color space, middle gray is equivalent to 46.6% brightness, which is rounded to 50% in our algorithm.

When a response function is recovered, Debevec and Malik [1997] introduce the constraint that the logarithm of the luminance for 50% brightness is zero. In order to convert the logarithm of ambient luminance to zero, the calibrated luminance

$$L'(x,y) = \frac{L(x,y)}{L_{av}} \tag{2}$$

can be used. The inverse of the response function g is a mapping from the natural logarithm of exposure to the display pixel values $L_d(x, y)$. The exposure can be computed as the product of the calibrated luminance and the exposure time Δt. The operation of tak-

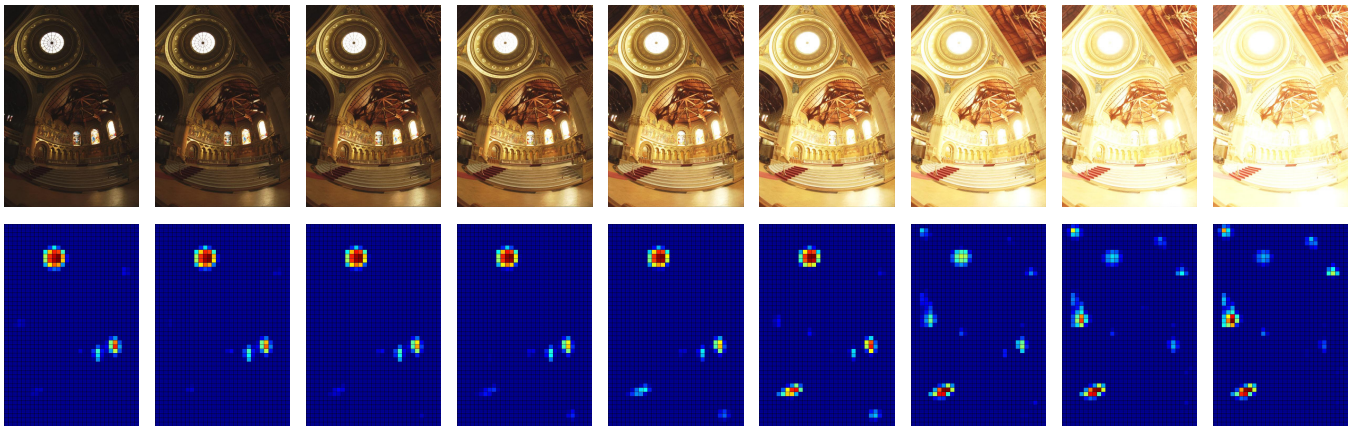

Figure 2: *Virtual photographs (upper row) and their corresponding saliency maps (lower row). The HDR image is due to Paul Debevec,* © *1997 ACM.*

ing virtual photographs is thus

$$L_d(x,y) = g^{-1}\left(\ln\left(L'(x,y)\Delta t\right)\right) \tag{3}$$

which is applied to each RGB color channel separately. Choosing $\Delta t = 1$ is akin to the automatic exposure control in a real camera.

A comparison between real and virtual photographs is shown in Figure 1. The real photographs have been used both to generate an HDR image and to recover the response function of the camera that the photographs were taken with. The same response function has been used to generate the virtual photographs from the HDR image. The virtual photographs look very similar to the real originals. We do note some very minor differences, since the exposure times of virtual photographs are based on ambient luminance and do not completely match those of the real photographs. However, the differences are too small to significantly affect the saliency analysis. Figure 5 below presents evidence that the performance of our algorithm is robust with regard to the exact choice of response function.

3.2 Saliency Analysis

In order to preserve details of HDR images while reducing artifacts that may be introduced by under- or over-exposed images, we calculate a sequence of nine virtual photographs from the HDR image with exposure times $\Delta t \in \{1/15, 1/8, 1/4, 1/2, 1, 2, 4, 8, 15\}$. We adopt the method of Itti and Koch [2000] to calculate the saliency map of each image in the sequence of virtual photographs, but other salience detection operators could be used instead.

Then the saliency maps are combined into the saliency map of the HDR image. One difficulty in combining different saliency maps is the signal-to-noise ratio problem identified by Itti and Koch [2000], which means that some salient objects may be weakened or entirely lost during combination. We solve this problem by adopting the spatial competition scheme used by Itti and Koch [2000] for noise reduction in the feature maps, which is realized employing a two-dimensional difference-of-Gaussians filter. The salient locations in each saliency map can be excited with counteraction triggered by the inhibition from the surrounding regions.

After the within-feature competition, the saliency maps are combined into the saliency map for the HDR image by computing their weighted average, where weights decrease with increasing degrees of over- and under-exposure. A similar strategy is used when producing HDR images from multiple LDR images with different exposures. Debevec and Malik [1997] make use of a simple hat function for the weights while Mann and Picard [1995] use the derivative of the response curve as the weighting function. In this paper, we (somewhat arbitrarily) employ $\phi_\sigma(\log_2 \Delta t)$ with $\phi_\sigma(x) = \exp(-(x/\sigma)^2/2)/\sqrt{2\pi\sigma^2}$ and $\sigma = 3$ to weight the saliency map obtained from the virtual photograph with exposure time Δt.

The focus of attention will be directed to the regions with the highest values in the saliency map. To that end, we utilize a winner-take-all (WTA) neural network [Koch and Ullman 1985] to determine the saliency locations and a fixed-size circle to represent the focus of attention. Inspired by biological neural networks, WTA networks work well to simulate the selection mechanism of the human brain [Frintrop et al. 2010].

A sequence of virtual photographs and their saliency maps are shown in Figure 2. It is apparent that some salient features are present only in a subset of the virtual photographs, and therefore only in some parts of the dynamic range. Prior methods fail to capture this. The final results from combining the salience maps of the virtual photographs as well as of the real photographs using the algorithm described above are shown in Figure 3. Here as well as in all of the following figures, the HDR image is tone mapped using the photographic tone mapping operator by Reinhard et al. [2002] for display purposes. It can be seen that the virtual photographs contribute to a similar saliency map and yield the same locations of visual attention as the sequence of real photographs.

From our experience, our approach is rather insensitive to the choice of response function. Two examples of response functions, which we have recovered from sequences of real photographs using the approach of Debevec and Malik [1997], are shown in Figure 4. Figure 5 shows saliency maps computed using those functions. The maps are almost identical, and the minor differences do not affect the further analysis, such as region selection for visual attention. In all of what follows, response function A is used.

4 Results and Discussion

We have tested our algorithm on HDR images with a broad range of content. Three examples are shown in Figure 6. In each case, our method is able to produce reliable and comprehensive results. The targets that differ from the surrounding environment by their unique colors, intensities, or orientations are accurately recognized. We have made similar observations in further test cases.

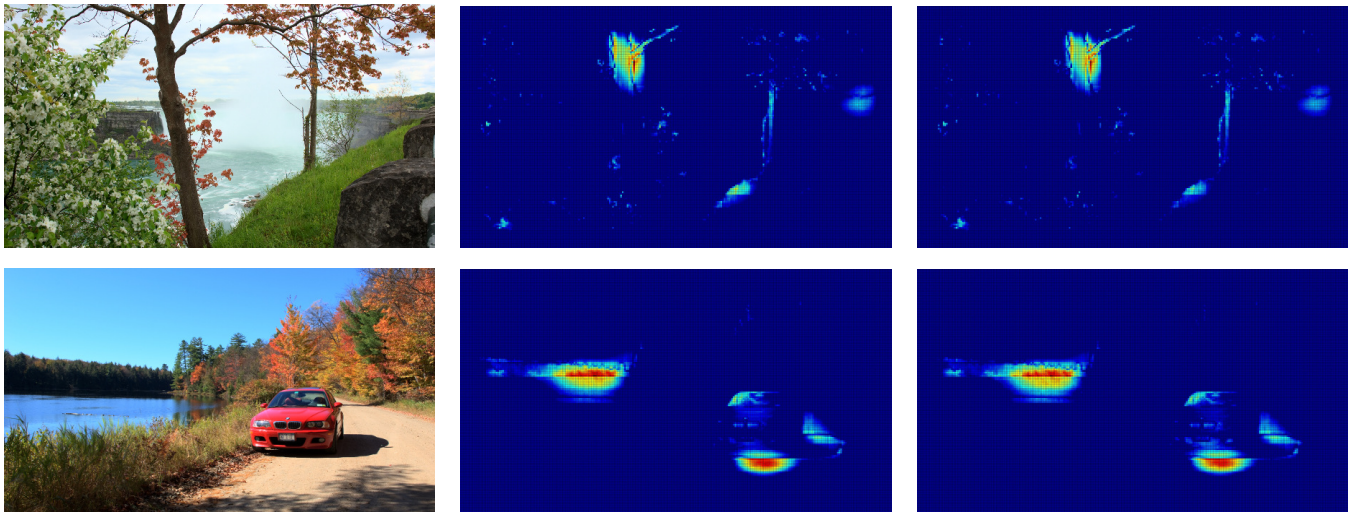

Figure 5: *Results of saliency analysis with different response functions. First column: the HDR images. Second column: the saliency maps computed with response function A. Third column: the saliency maps computed with response function B. The HDR images are from* Mark Fairchild's HDR Photographic Survey*, © 2006-2007 Mark D. Fairchild.*

A comparison with results generated by applying the method of Itti and Koch [2000] (which is targeted to LDR images) directly to the HDR image as well as those obtained using the algorithm of Brémond et al. [2012] is shown in Figures 7, 8, and 9. Compared with the earlier approaches, our method addresses the problem of HDR content loss that occurs due to brightness compression, and therefore consistently leads to better performance for detecting the salient regions of HDR images. As illustrated in Figure 7, if the luminance of the HDR image is compressed before the saliency analysis (as is the case with the other methods), it is hard for the visual models to analyze details in the dark areas of the image. In this example, only the sun and its reflection on the water surface are detected by the other approaches, while the salient targets in other regions, such as the mountains, the sailboat, and the house, are neglected. The same pattern occurs in Figure 8. The previously cited methods fail to capture the targets in the background, which is darker than the foreground. Another example is given in Figure 9. Our method detects the trees, which clearly are eye-catching, while the other methods do not. We hypothesize that this is because the details in relatively dark areas are weakened or even entirely lost as a result of significant changes in contrast.

5 Conclusion

A number of computational attention systems have been proposed for LDR images. However, relatively poor results have been observed when applying some of those systems for saliency analysis of HDR images since lightness reduction will lead to significant loss of details. To preserve details of HDR images and incorporate them into the saliency analysis, a virtual photograph based method is presented in this paper. The approach utilizes a sequence of virtual photographs rather than a single image for revealing HDR content. Our method reliably characterizes regions of visual attention for HDR images and has a wide variety of potential applications, such as HDR image or video coding, computational aesthetics, non-photorealistic rendering, evaluation of tone mapping operators, and saliency-based tone mapping algorithms. Moreover, the virtual photograph technique presented in our paper opens new avenues for analyzing HDR content, and it may be beneficial for computational photography as well. In future work, we plan to more formally analyze the relative performance of our approach by using eye tracking data.

Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the GRAND Network of Centres of Excellence.

References

AGRAWAL, A., RASKAR, R., NAYAR, S. K., AND LI, Y. 2005. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics 24*, 3, 828–835.

AVIDAN, S., AND SHAMIR, A. 2007. Seam carving for content-aware image resizing. *ACM Transactions on Graphics 26*, 3, 10:1–10:10.

BRÉMOND, R., PETIT, J., AND TAREL, J. P. 2012. Saliency maps of high dynamic range images. In *Trends and Topics in Computer Vision — ECCV 2010 Workshops*. Springer Verlag, 118–130.

BROWN, B. 2011. *Cinematography: Theory and Practice — Image Making for Cinematographers and Directors*. Focal Press.

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques — SIGGRAPH '97*, ACM Press, 369–378.

DONG, L., LU, S., AND JIN, X. 2013. Real-time image-based Chinese ink painting rendering. *Multimedia Tools and Applications*, to appear.

FRINTROP, S., ROME, E., AND CHRISTENSEN, H. I. 2010. Computational visual attention systems and their cognitive foundation: A survey. *ACM Transactions on Applied Perception 7*, 1, 6:1–6:39.

FRINTROP, S. 2006. *VOCUS: A visual attention system for object detection and goal-directed search*. Springer Verlag.

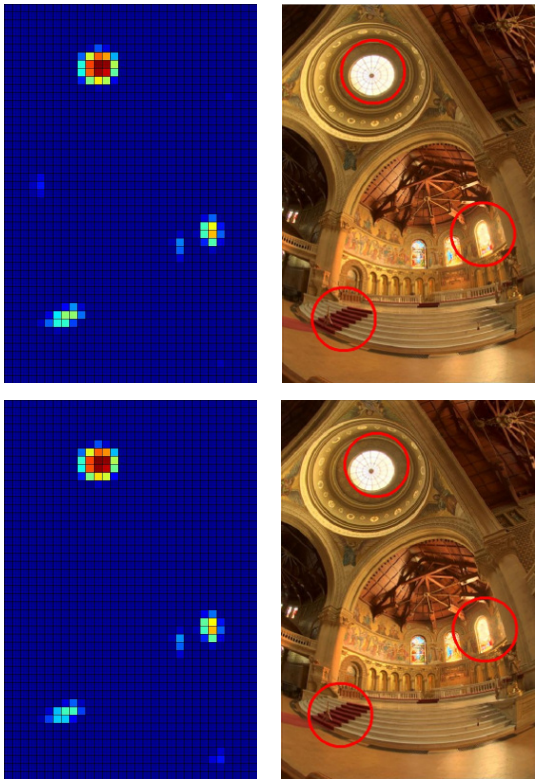

Figure 3: *Comparison between the results from virtual and real photographs. Upper row: saliency map and regions of visual attention (red circles) from virtual photographs. Lower row: saliency map and regions of visual attention from real photographs. The HDR image is due to Paul Debevec, © 1997 ACM.*

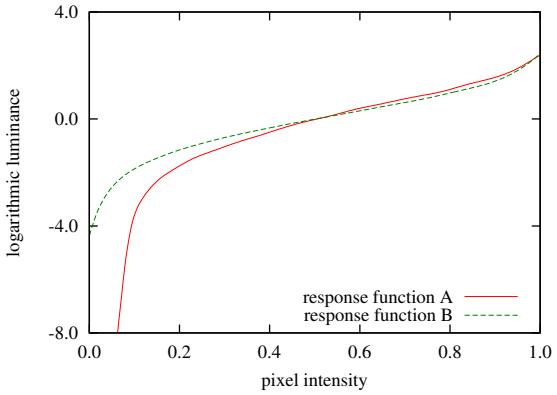

Figure 4: *Two response functions mapping pixel intensities to luminance values.*

ITTI, L., AND KOCH, C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research 40*, 10-12, 1489–1506.

ITTI, L., KOCH, C., AND NIEBUR, E. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 20*, 11, 1254–1259.

ITTI, L. 2004. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing 13*, 10, 1304–1318.

KOCH, C., AND ULLMAN, S. 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology 4*, 4, 219–227.

LEE, H., SEO, S., RYOO, S., AHN, K., AND YOON, K. 2013. A multi-level depiction method for painterly rendering based on visual perception cue. *Multimedia Tools and Applications 64*, 2, 277–292.

LIN, W., AND YAN, Z. 2011. Attention-based high dynamic range imaging. *The Visual Computer 27*, 6, 717–727.

MANN, S., AND PICARD, R. W. 1995. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *48th Annual IS&T Conference*, Society for Imaging Science and Technology, 422–428.

MOHAN, A., PAPAGEORGIOU, C., AND POGGIO, T. 2001. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 4, 349–361.

NARWARIA, M., PERREIRA DA SILVA, M., LE CALLET, P., AND PÉPION, R. 2012. Effect of tone mapping operators on visual attention deployment. In *SPIE Proceedings Vol. 8499 — Applications of Digital Image Processing XXXV*. International Society for Optics and Photonics.

REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics 21*, 3, 267–276.

RICHARDT, C. 2008. Flash-exposure high dynamic range imaging: Virtual photograph and depth-compensating flash. Tech. Rep. UCAM-CL-TR-712, University of Cambridge.

TREISMAN, A. M., AND GELADE, G. 1980. A feature-integration theory of attention. *Cognitive Psychology 12*, 1, 97–136.

VIOLA, P., AND JONES, M. 2004. Robust real-time face detection. *International Journal of Computer Vision 57*, 2, 137–154.

WALTHER, D., AND KOCH, C. 2006. Modeling attention to salient proto-objects. *Neural Networks 19*, 2, 1395–1407.

WANG, Y. S., TAI, C. L., SORKINE, O., AND LEE, T. Y. 2008. Optimized scale-and-stretch for image resizing. *ACM Transactions on Graphics 27*, 5, 118:1–118:8.

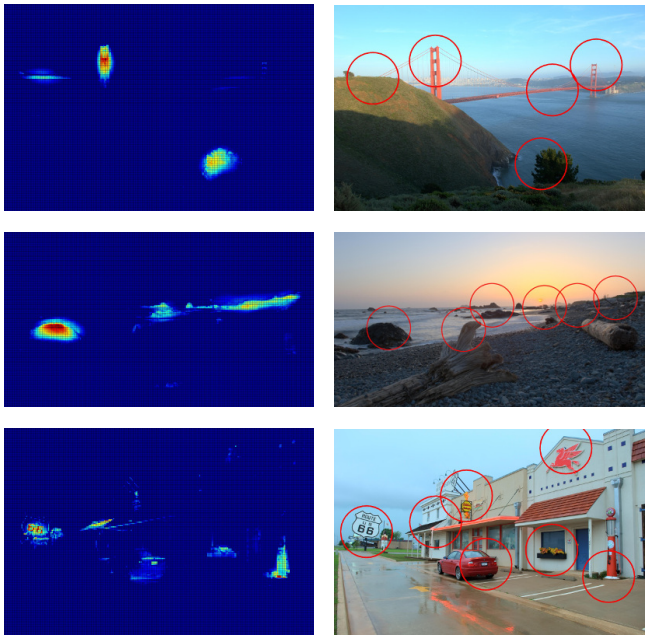

Figure 6: *Results of saliency analysis for three test images. Left hand column: saliency maps. Right hand column: locations of visual attention (red circles). The HDR images are from* Mark Fairchild's HDR Photographic Survey, © *2006-2007 Mark D. Fairchild.*

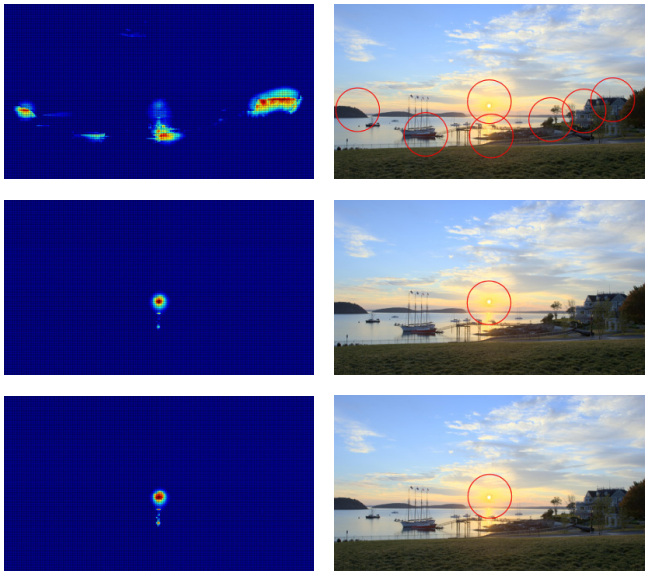

Figure 7: *Comparison with other approaches. From top to bottom, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [2000], and the method of Brémond et al. [2012]. The HDR image is from* Mark Fairchild's HDR Photographic Survey, © *2006-2007 Mark D. Fairchild.*

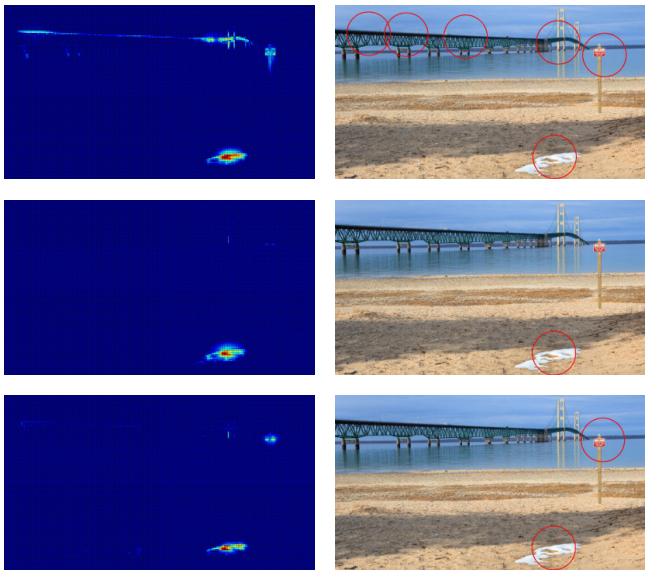

Figure 8: *Comparison with other approaches. From top to bottom, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [2000], and the method of Brémond et al. [2012]. The HDR image is from* Mark Fairchild's HDR Photographic Survey, © *2006-2007 Mark D. Fairchild.*

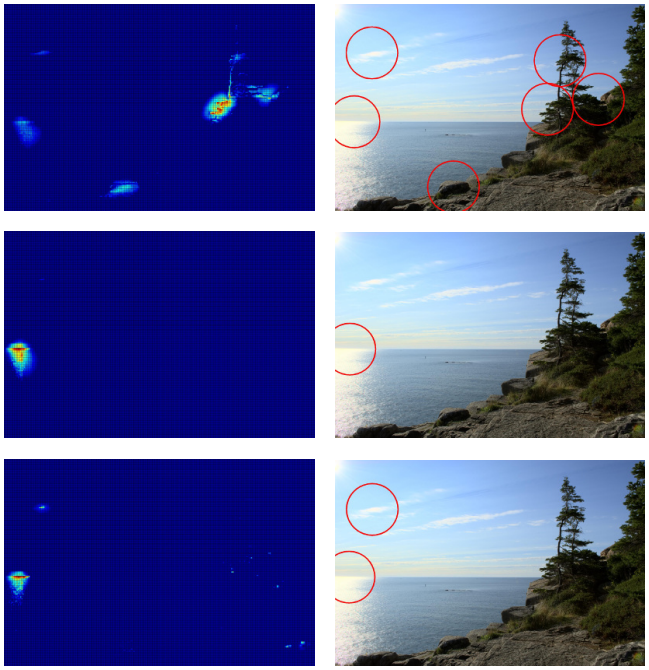

Figure 9: *Comparison with other approaches. From top to bottom, saliency maps and regions of visual attention (red circles) produced by our method, the method of Itti and Koch [2000], and the method of Brémond et al. [2012]. The HDR image is from* Mark Fairchild's HDR Photographic Survey, © *2006-2007 Mark D. Fairchild.*