

Nebula: Live Dynamic Projection Mapping via Object Saliency

Sara Greenberg, Audrey G. Chung, and Alexander Wong
 Systems Design Engineering,
 University of Waterloo
 200 University Ave W, Waterloo, ON N2L 3G1
 smgreenb@uwaterloo.ca

Abstract

Saliency algorithms provide a quantitative measurement of importance for points or regions of an image. This can be valuable to digital artists who wish to produce works that build layers of abstraction or augmentation onto an input image or video frame. A saliency metric developed for illumination-robust object saliency can be used as a framework for artistic mapping, and one result is a live, multi-disciplinary projection piece. Nebula is an acrobatic performance piece in which an aerial dancer's movement is augmented by projected starscapes created in real time using a depth camera live feed and the described framework.

Introduction

Abstracting or augmenting a raw image is a common practice in various artistic disciplines, such as photo manipulation and projection mapping. Often, the image features of interest are identified manually using programs such as Isadora [1] or Resolume Arena [2]. Objects can also be tracked automatically. Simple implementations include thresholding an image for a single color, e.g., tracking a red ball in a field. Objects can also be tracked by searching for sets of known mathematical features, which is what SURF [3] does. While these methods work well when the object being tracked is static and has predictable features, they fail when the tracked image is not easily defined [4]. Computational methods do exist for the automatic detection of salient objects or regions within an image [5].

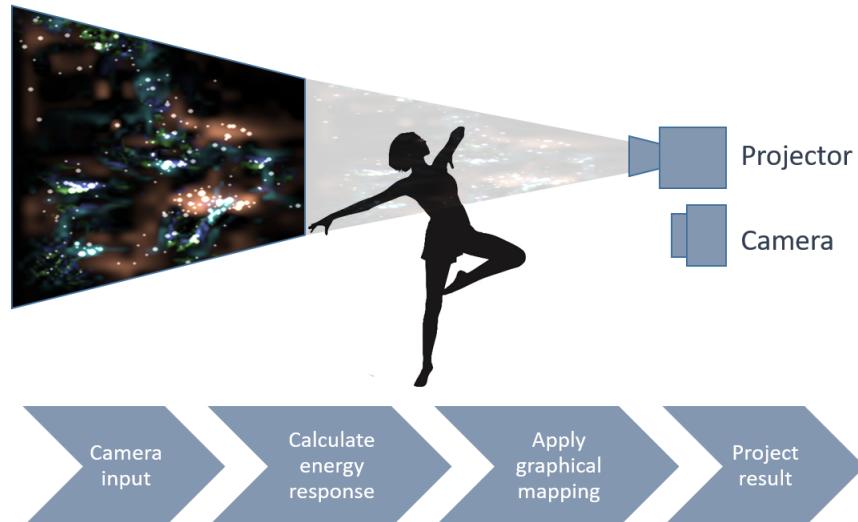


Figure 1: Example use of multi-scale saliency and the aggregate energy response (described below) for live projection-augmented performance.

One such algorithm developed by the Vision and Image Processing Research Group is the Texture-Illumination Guided Global Energy Response (TIGGER) [6]. The goal of TIGGER is to identify the most salient objects within an image by identifying locations of interest at several image scales, and then combining this multi-scale information to get a global saliency measure for each pixel within the image. In this case, salient pixels in an image are ones that are different from their neighboring pixels where many sizes of neighborhoods are considered. To correct for lighting inconsistencies, TIGGER also re-balances the texture and illumination information of the image.

Because TIGGER produces a set of multi-scaled responses during execution, it provides two opportunities for mapping: first, each of the scale responses is a quantitative measurement that is sensitive to a specific range of object size, and second, the aggregate response for the entire image. These components provide a rich interpretation of the “important” parts of an image, which can be mapped to creative imagery. This algorithm became the basis for the performance piece *Nebula* by Sara Greenberg, in which a live, dynamic projection mapping is displayed over aerial hoop choreography. This piece examines the human condition and our place in the universe, using Carl Sagan’s reflections on the Pale Blue Dot photograph [7] as an audio track while a variation of TIGGER produces a moving starscape in response to Sara’s location on stage.

The flow of this system is described in Fig. 1, where a performance is augmented by capturing the scene using a camera, processing the image to acquire the energy response and saliency metrics, applying a form of graphical mapping, and finally, re-projecting the result back onto the scene. In *Nebula*, the projection and camera are aligned to create the illusion that the performer is the nexus of a star cloud. A depth camera (which acquires a grayscale image using an infrared emitter and sensor) is used in order to avoid creating a feedback loop between the input image and the projected image.

About the Aggregate Energy Response

TIGGER functions, as shown in Fig. 2, by first separating the texture and illumination information of an image. Then, the image is scaled to multiple sizes, and an energy response (a mathematical evaluation of saliency) is determined for each pixel in each layer, or scaled image. This allows for the detection of salient features at a variety of scales. The results of each layer are re-sized to the original scale and combined to acquire an aggregate energy response (AER). The original image is divided into segments using K-means clustering, and a region adjacency graph (RAG) classifies all segments within the image according to the saliency map.

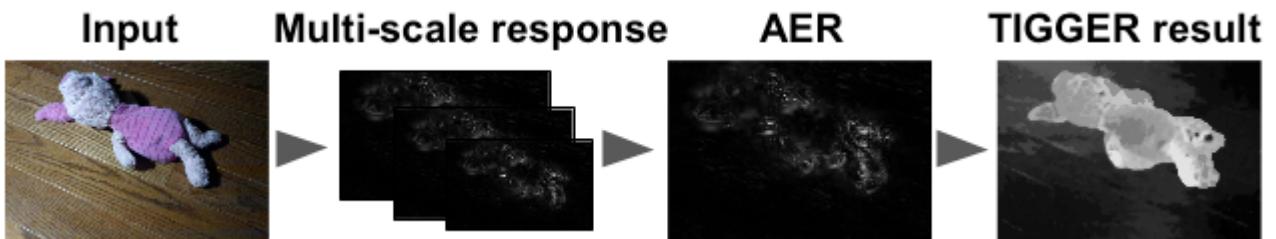


Figure 2: TIGGER process, in which energy responses are calculated on multiple image scales and combined to form an aggregate energy response (AER), which is then applied to a region adjacency graph.

The texture-illumination decoupling can be bypassed, either to achieve more interesting results or because the source image is not affected by lighting changes. The remaining functionality is the global energy response, and because the layering of multiple sub-responses is of interest, it will be referred to here as the aggregate energy response (AER). This response is accomplished by evaluating the changes in pixel values

within all scaled images. At each scale factor λ , the saliency metric is computed using the Hessian matrix, Φ , which calculates a matrix of second-order partial derivatives for each pixel, \bar{q} , of an image, I , as

$$\Phi(\bar{q}, \lambda) = \begin{bmatrix} (\Delta_{x,\lambda} I(\bar{q}))^2 & \Delta_{x,\lambda} I(\bar{q}) \Delta_{y,\lambda} I(\bar{q}) \\ \Delta_{y,\lambda} I(\bar{q}) \Delta_{x,\lambda} I(\bar{q}) & (\Delta_{y,\lambda} I(\bar{q}))^2 \end{bmatrix} \quad (1)$$

where Δ_x and Δ_y are gradients in the x and y directions, respectively. The saliency s at scale λ of pixel \bar{q} is a scalar indicating how different the pixel in question is from its surroundings (Eq. 2). The global, aggregate response is a summation of all scales (Eq. 3).

$$s(\bar{q}, \lambda) = \frac{\det(\Phi(\bar{q}, \lambda))}{\text{trace}(\Phi(\bar{q}, \lambda))} \quad (2) \quad \Theta(\bar{q}) = \sum_{\lambda \in \Lambda} \frac{\det(\Phi_\tau(\bar{q}, \lambda))}{\text{trace}(\Phi_\tau(\bar{q}, \lambda))}. \quad (3)$$

Making Art With It

The development of TIGGER was coincident with a research project involving human body tracking using the Microsoft Kinect. The Kinect is a device containing both an RGB camera and a depth camera that functions using infrared light. As an experiment, TIGGER was applied to various depth images (such as Fig. 3a). The texture-illumination decoupling is unnecessary for these images, and when the process is halted before the classification of all segments using the RAG model, the result is a cloudy and abstract form of the original image, i.e., the aggregate energy response (AER). Each layer of the result can be given a color by randomly selecting RGB values, as shown in Fig. 3b. Since the original algorithm is optimized for RGB images, the results for depth images appear to pick up areas with more features such as edges and corners.

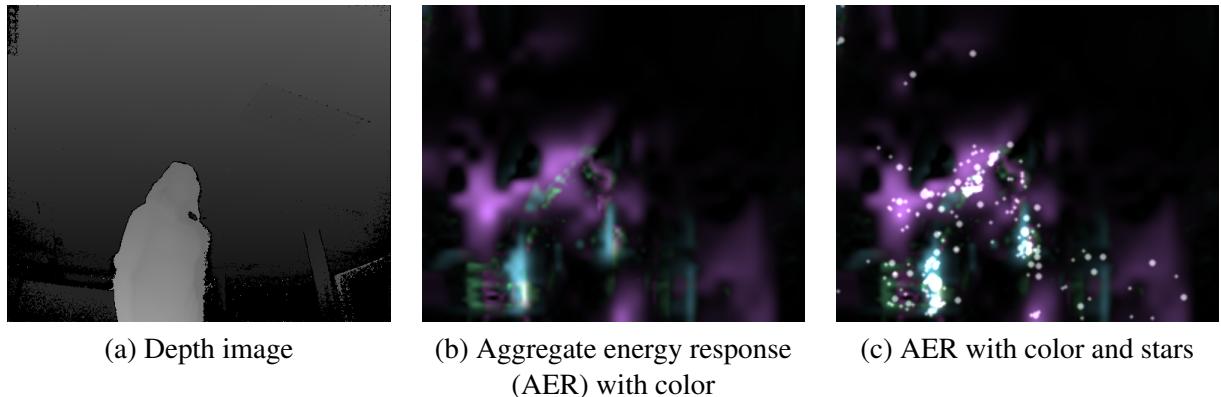


Figure 3: Example of mapping color to each scaled response, and white circles with radii proportional to the aggregate response value at each pixel.

Another method for graphical mapping is to apply some creative content to parts of the image within a saliency threshold. For example, the appearance of stars can be created by randomly sampling pixels from the image and selecting only pixels with an energy response above some threshold as described by the acceptance-rejection sampling in Algorithm 1. We can then apply a white circle with radii proportional to the saliency value at each selected pixel, centered at the pixel location (Fig. 3c).

Graphical mappings are not limited to colors and circles. The result of the energy response is highly dependent on the number and size of the scale factors (Eqs. 1, 2, and 3), as well as the number of pixels over which the Hessian matrix calculates its second-order derivatives. While these parameters are tuned to improve the object detection performance of the original algorithm, these parameters can be adjusted according to the artist's vision in creative applications.

Algorithm 1: Acceptance-rejection sampling for a pixel \bar{q} .

```

count = 0 ;
while count < desired samples do
|   if  $\Theta(\bar{q}) > \text{threshold}$  and  $\Theta(\bar{q}) > \text{rand}$  then
|   |   accept  $\bar{q}$  ;
|   |   count++ ;

```

Nebula

The performer of Nebula, Sara, often finds herself living two lives: as an academic researcher, and as a performer. This image processing artistic framework has allowed her to fuse her passions into a cohesive performance. Nebula is an aerial hoop piece examining the human condition and our place in the universe, using Carl Sagan's reflections on the Pale Blue Dot photograph [7] as a basis. Using the aggregate energy response, a live starscape is created from a depth camera feed of her location on stage. The result (Fig. 4) is an image that maintains some of the shape and features of the input depth image, but is abstracted to resemble a nebula, galaxy, or other astronomical object.



Figure 4: *Nebula*, performed at Launch in Toronto, Ontario, April 2016.

The described algorithm for Nebula is calculated in under one second, and is applied to a live video sequence with some delay. The frames are interpolated to improve smoothness, and projected onto Sara's acrobatic performance. This piece has been performed at two public showcases in Ontario, Canada.

References

- [1] S. deLahunta, "Isadora almost out of beta: tracing the development of a new software tool for performing artists," *International Journal of Performance Arts and Digital Media*, vol. 1, no. 1, pp. 31–46, 2005.
- [2] H. Jung, J. Lee, H.-J. Choi, and H. Kim, "Real-time djing + vjing with interactive elements," *Contemporary Engineering Sciences*, pp. 1321–1327, 2014.
- [3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [4] O. Miksik and K. Mikolajczyk, "Evaluation of local detectors and descriptors for fast feature matching," in *International Conference on Pattern Recognition (ICPR)*. IEEE, 2012, pp. 2681–2684.
- [5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 10, pp. 1915–1926, 2012.
- [6] S. Greenberg, A. G. Chung, B. Chwyl, and A. Wong, "TIGGER: A texture-illumination guided global energy response model for illumination robust object saliency," in *Conference on Computer and Robot Vision*, 2016.
- [7] C. Sagan, *Pale Blue Dot: A Vision of the Human Future in Space*. Random House Digital, Inc., 1997.