ICam: Maximizes Viewers' Attention on Intended Objects

Rajarshi Pal, Pabitra Mitra, and Jayanta Mukhopadhyay

Department of Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India
{rajarshi,pabitra,jay}@cse.iitkgp.ernet.in

Abstract. Continuous inventions of modern cameras are making the act of good photography easier. This paper proposes a methodology that makes a camera even smarter by converting the objects of interest to the photographer more salient. This results in maximization of viewers' attention on intended objects. The captured photo can be taken as a media of communicating photographer's mind to the potential viewers- in that sense; the proposed methodology reduces the communication gap between them.

Keywords: Intelligent camera, Visual attention.

1 Introduction

At the early stage of evolution, camera research was mainly confined in increasing resolution. Now when moderately high resolution cameras are coming within affordable prices, the focus of evolution has shifted - towards smarter cameras. In [1], an imaging device is proposed using multiple beam splitters to capture a scene through the same optical axis but with different parameters. In [2], a multi-aperture camera is designed and various applications of using multiple aperture settings is depicted. 'The moment camera' in [3], has added another new feather in the area of computational photography. In [4] a reprogramming of auto-gain mechanism is proposed to capture high dynamic range images/videos. LAFCam [5] and StartleCam [6] are two more examples of creative solutions to meet the demand of passionate photographers. And, more interestingly, now-a-days cameras (Canon PowerShot series, Nicon's CoolPix seires, Olympus' μ series, Fujifilm's FinePix series, Sony's Cybar-shot series) have face detection capability to better focus on human faces present in the filed of view of camera.

One of the basic composition rules of photography is to eliminate distracting elements from the view of the scene to be taken. When people go for an outing, they try their hands at camera, and have a desire to excel as photographers. Sometimes it happen that the photographer wanted to capture something, but some other things (which the photographer did not notice in a hurry while taking the shot) appear to draw more attention from viewers. This is because a camera captures everything in its field-of-view. It may not be always possible for the photographer to minutely scan the LCD panel while composing. In this paper,

Y.-M.R. Huang et al. (Eds.): PCM 2008, LNCS 5353, pp. 821–824, 2008.

we express the idea of having an intelligent camera (ICam) that uses selective visual attention models to estimate which portions of the photograph will quickly attract viewers' gaze. If a distracting obejct is found, it automatically readjusts its parameters to compose the shot eliminating that distracting object. Thus, it maximizes the possibility of drawing viewers' attention to those portions of the picture where the photographer wants them to attend. ICam uses selective visual attention models to estimate which portions of the photograph will quickly attract viewers' gaze. Then there will be no need to care for these things by the photographer and everyone can become a good photographer.

The outline of this paper is as follows: Section 2 briefly discusses what selective visual attention is. Section 3 describes the concept of ICam. Section 4 shows some results obtained from our simulation and finally we conclude in Section 5.

2 Selective Visual Attention

To predict the portions of the image that will draw viewers' attention we have used selective attention models. These models are applied on the image to obtain visually salient locations. In our prototype of ICam, we have used the model proposed in [7]. Salient locations are extracted based on multi-scaled representation of the following low-level features: red-green and blue-yellow opponent color channels, intensity, and orientation at four directions ($0°$, $45°$, $90°$ and $135°$).

3 Description

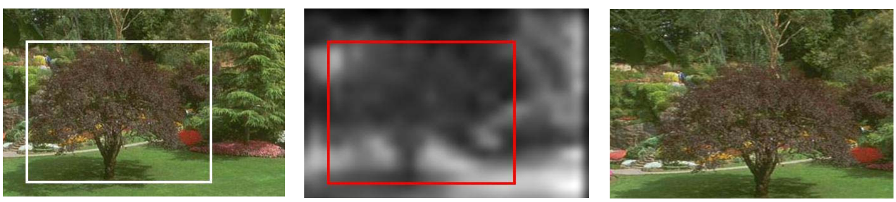
Photographer sees through the LCD panel and informs the camera about the objects of interest to him. To ease the implementation of our prototype, interested objects are communicated by pointing four corners of the minimum rectangle that encompasses the objects of interest. This rectangle may not be the minimum rectangle in true sense, as it is very fast rough estimation made by the photographer.

When the sensor is exposed to light, saliency map of the image formed in the sensor is estimated. It is checked whether other objects are more salient than the objects in which the photographer is interested. The desired condition is that interesting objects must be salient to quickly grab viewers' attention. When other objects are more salient than the objects of interest, the camera lens is to be positioned in such a way that its view avoids the non-interesting but salient objects. In our designed prototype, the focus of the lens can shift at four basic directions (up, down, left and right) and a few possible combinations of these four directions (up+left, up+right, down+left, down+right). After analyzing the positions of the non-interesting but salient objects, the focus of the lens is shifted to a particular direction. Interesting fact is that the focus cannot be shifted at left and right directions simultaneously (similar is the case for up and down). So, if the non-interesting but salient objects appear in opposite directions with respect to the important objects, the lens is zoomed in to avoid those non-interesting objects.

Once the focus is shifted, a slight variation of the previously captured scene will fall on the sensors. Again, saliency map of this new image will be estimated and the saliency of interesting/non-interesting objects will be compared and the focus of the lens will be shifted if needed. This process will continue until the sensor captures a scene where interesting objects are more salient than the non-interesting objects. At the end, the captured scene will be stored as the photograph of the desired object taken by the photographer.

4 Results

Figure 1(a) shows the initial image captured on the sensor while the photographer wanted to take the picture of the tree as indicated by the white-bordered rectangular area imposed on the figure. Figure 1(b) shows the saliency map of the first image of the scene captured on the sensor. The area inside the red rectangle imposed on the figure indicates the interesting area. And in the saliency map the more brighter an area means more salient the area is. Clearly there are some white regions outside the rectangle in figure 1(b). This means the focus



(a) (b) (c)

Fig. 1. (a) Initial image captured by the camera, (b) Saliency map of the initial image, (c) Final version of image captured by ICam

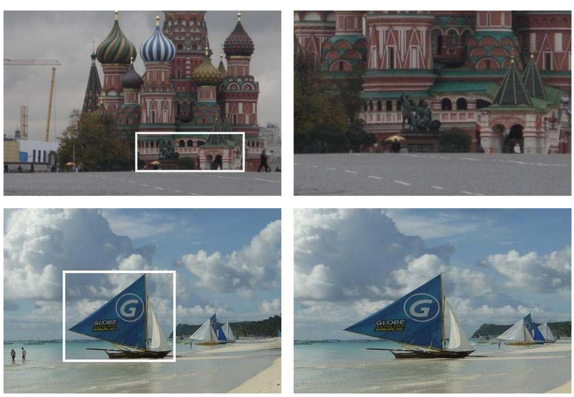


Fig. 2. Left: Initial images, Right: Final versions captured by ICam

of the lens has to be shifted to avoid those areas. Figure 1(c) shows the correct image taken by ICam after a few adustments in its focus. Figure 2 shows some more examples of initial images formed on the sensor along with the areas in which the photographer was interested and the final image captured in ICam.

5 Conclusion and Future Work

ICam is designed to ease the act of good photography. Rendering the objects of interest as more salient, ICam succeeds to attract viewers' gaze on the intended objects. Our proposed methodology can also be blended with other interesting features that make the cameras handy to use.

Modeling the important objects using a rectangle is a drawback of our prototype. The proposed prototype works based on the assumption that everything inside the rectangle is of importance to the photographer. But this shouldn't be the case always. Figure 1 demonstrates such a case. While taking the picture, photographer's real interest was on the tree. Figure 1(b) shows that the tree is less salient but the grasses underneath the branches and leaves of the tree are salient. In order to encompass the entire tree (along with its stem) with the rectangle, the photographer has also included those salient areas inside the rectangle. In future, we plan to include the arbitrariness in the shape of the interesting objects.

References

1. McGuire, M., Matusik, W., Pfister, H., Chen, B., Hughes, J.F., Nayar, S.K.: Optical Splitting Trees for High-Precision Monocular Imaging. IEEE Computer Graphics and Applications 27(2), 32–42 (2007)
2. Green, P., Sun, W., Matusik, W., Durand, F.: Multi-Aperture Photography. ACM Transactions on Graphics 26(3) (2007)
3. Cohen, M.F., Szeliski, R.: The Moment Camera. Computer 39(8), 40–45 (2006)
4. Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High Dynamic Range Video. ACM Transactions on Graphics 22(3), 319–325 (2003)
5. Lockerd, A., Mueller, F.: LAFCam - Leveraging Affecting Feedback Camcorder. In: CHI 2002 extended abstracts on human factors in computing systems, pp. 574–575 (2002)
6. Healey, J., Picard, R.W.: StartleCam: A Cybernetic Wearable Camera. In: 2nd International Symposium on Wearable Computers, pp. 42–49 (1998)
7. Itti, L., Koch, C., Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)