Photo Quality Assessment based on a Focusing Map to Consider Shallow Depth of Field

Dong-Sung Ryu, Sun-Young Park, Hwan-Gue Cho
Dept. of Computer Science and
Engineering, Pusan National
University, Geumjeong-gu BUSAN,
South KOREA, 609-735
dsryu99,parksy,hgcho@pusan.ac.kr

ABSTRACT

Proliferation and advances in digital cameras encourage people to take many photos. However, the number of photos that people can access is increasing exponentially. Good quality photo selection is becoming burdensome. In this paper, we propose a novel method to evaluate photo quality considering DoF (Depth of Field) based on a focusing map. The focusing map is a form of saliency map classified into four levels based on the spatial distribution of Canny edges. We implemented it in a CUDA environment to improve the speed of focusing map generation. In order to evaluate our method, we tested our feature on the four classified 206 photos; then, we compare our method to a photo set manually classified by a user. The proposed measure efficiently assesses the photos with DoF. Especially, the expert group who used DSLR camera agreed that our photo assessment measure is useful.

Keywords: digital photo, photo assessment, depth of field.

1 INTRODUCTION

Generally, high quality photos satisfy three principles: 1) a clear topic, 2) a focus of attention on the subject, and 3) the removal of objects that distract attention from the subject [4]. DSLR camera users control the camera parameters, for example, aperture and shutter speed, to take good quality photos. The assessment of these photos depends on how to arrange and present photo subject clearly. In the case of DoF photos,it is very easy to know which object is its subject due to the their out-of-focussed background region. Therefore, it is important for DSLR camera users to evaluate DoF features in selecting good quality photos. In this paper, we define high quality photos as those that have shallow DoF features. Figure 1 shows the characteristics of good quality photo with DoF. The deep DoF photo is insufficient for a good quality photo, since its topic (the book) in Figure 1 (a) is not presented clearly (The region of the book is blurred).

Most studies related to good quality photo evaluation proposed combined measures of contrast, blur, and hue count to evaluate image degradation caused by noise, distortion, and compression artifacts. These studies are efficient in distinguishing defective (i.e. blurred) photos [2, 6]. In contrast to these works, we proposed a DoF measure based on the focusing map. The focusing map is a form of saliency map with four classification levels based on the spatial distribution of Canny edges. We calculated the total weight of all Canny edge pixels allocating four weight values corresponding to the focusing level.

(a) (b)

Figure 1: Examples of good quality photos considering difference in DoF. (a) A deep DoF photo. (b) A shallow DoF photo.

2 RELATED WORK

Photo assessment studies measure color contrast and blur caused by camera shaking, overexposure and misconfigured camera settings. Recent studies deal with visual features in photography more importantly than ever before. Ke *et al.* designs high level image features to measure perceptual differences [2], considering the spatial distribution of edges, color distribution, hue count, blur and contrast. They combined these features using Bayes rules. Datta summarized 56 features to consider aesthetics in photography [1]. These classifiers are built using support vector machines and classification trees. That work focuses on the relationship

WSCG 2011 conference proceedings
WSCG'2011, January 31 – February 3, 2011
Plzen, Czech Republic.

between emotions that pictures arouse in people, and their low-level content. Sun *et al.* proposed the RFA (Rate of Focused Attention) measurement based on a saliency map of computational visual attention model to consider human visual systems for photo assessment [5]. This model simulates the attention mechanism of a human visual system, most of which use a saliency map or a conspicuous map to describe how salient (interesting) a location in the visual field is. All these recent studies tried to find optimized visual features to mimic human perception in photo assessment.

Luo *et al.* designed a composition that depicts the organization of all graphical elements within a photo for professional photos [4]. They use a log-likelihood of derivatives with a blurring kernel of size $k \times k$ ($1 \leq k \leq 50$) to extract the subject region and to search the interest region of photos.

3 DEPTH OF FIELD PHOTO EVALUATION

Figure 2 depicts the framework of our method. It is important how a human detects a recognizable region in a photo. We first extract the Canny edge from the blurred photo to compute the focusing map to achieve this. The focusing map estimates the visual attention regions of a photo which consists of four regions. These visual regions are classified according to the count of Canny edge pixels in a designated mask. Then, we can calculate the score of the DoF photo comparing all extracted Canny edge pixels with the focusing map pixels.

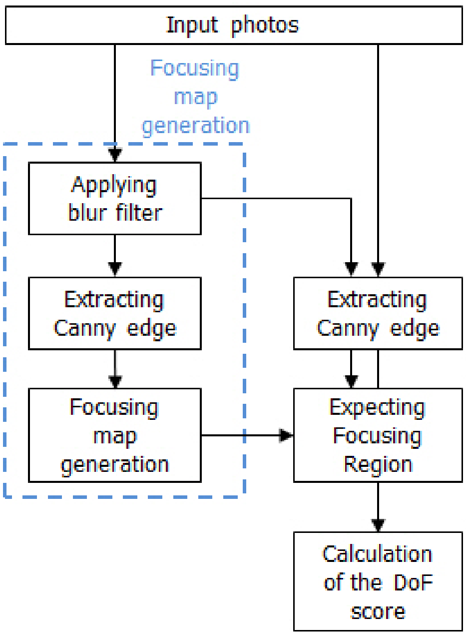

Figure 2: Framework of the proposed method.

3.1 Focusing Map Generation

Most studies use a saliency map to estimate the visual attention region, since it describes how salient a location in the image is. We used a similar concept of a saliency map [5]. However, we divide the photo regions into four classified pixel sets based on their intensity, R_k ($k \subseteq \{r, o, y, w\}$, each element depicting red, orange, yellow and white, respectively). These four levels are determined by the number of Canny edge pixels in a designated mask, m. Since Canny edge extraction considers the direction of pixels variation and double thresholds, human recognizable edges are extracted by using Canny edge detection algorithm. The description of the proposed focusing map generation is as follows. First, we apply the blur (Gaussian Filter) and Canny edge filter to the original images. This is to find the recognizable edge pixels. Then, we calculate the focusing value based on the count of Canny edge pixels in the designated mask (m) at each pixel. The mask size is $\sqrt{(w+h)/2}$, where w and h are the photo width and height, respectively. Third, in order to prevent the sequential Canny edge pixel counting from left top to right bottom image, all pixels in the focusing map are randomly shuffled. Then, they are sorted in the pixel value of the focusing map order. Finally, the sorted pixels are separated into four pixel groups (R_r, R_o, R_y, and R_w) whose sizes are 12.5%, 12.5%, 25%, and 50%, of the total number of pixels, respectively. These four pixel groups is determined by a naive experiment to be able to evaluate DoF photos in our 206 photo sets. This experiment is to maximize F-score of Section 4 Experiment. Note that each pixel on the image are divided into four pixel groups in the designated ratios. The role of these designated pixel subsets is to estimate which regions are focused in the entire image.

Let us explain an example of focusing map generation, as shown in Figure 3. We assumed the mask size is 3×3. Figure 3 (a) shows the extracted Canny edge pixels. We execute the mask operation every photo pixel to calculate the number of Canny edge pixels in the 3×3 mask. Thus, we can obtain the focusing map, as shown in Figure 3 (b).

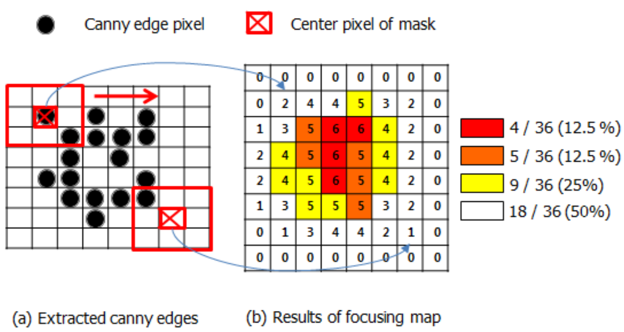

Figure 3: Example of the proposed focusing map.

The proposed focusing map is simple to implement. However, it has many iterative mask operations to scan all the pixels in a photo. Assume that we have a 1024×768 resolution photo, whose mask size is 100×100. It will need 7,864,320,000 pixel traverse operations (a time-consuming task). We applied CUDA (Compute Unified Device Architecture) to implement this procedure to improve the speed of focusing map generation. This GPU implementation improves system performance more than approximately $15 \sim 20$ times compared to CPU implementation. For example, the CPU implementation took about 33 *s* on average for the $1{,}024 \times 768$ resolution photos described in Table 1. With GPU implementation, it took approximately 1.9 *sec* on average.

3.2 Shallow and Deep Depth of Field Photos

Most professional photographers intentionally create a blurred area in the background region with a shallow DoF [3]. Therefore, a well-taken photo has a high possibility of having an intensive concentration of edges on an interesting object or region. We use the Canny edge detection algorithm to extract the human recognizable edges. The Canny edge considers the direction of pixel variation and double thresholds. We extract human recognizable edges from photos more readily than from other edge extraction methods. The main idea of our DoF photo assessment method is how to consider if the edges extracted from a photo image are in the in-focus or out-of-focus region. If edge pixels can be obtained from the in-focus region but no edge pixels can be obtained from the out-of-focus region, simultaneously, the photo is a shallow DoF photo. Figure 4 shows the process of the DoF quality evaluation method to consider an edge distribution that depicts the level of focus. We compute the number of the Canny edge pixels in the mask m corresponding to the level of the focusing map. The weight values for focusing map, w_r, w_o, w_y, and w_w are *5*, *1*, *-1*, *-5*, respectively. Our proposed focusing value $f(p)$ for each pixel p is defined by :

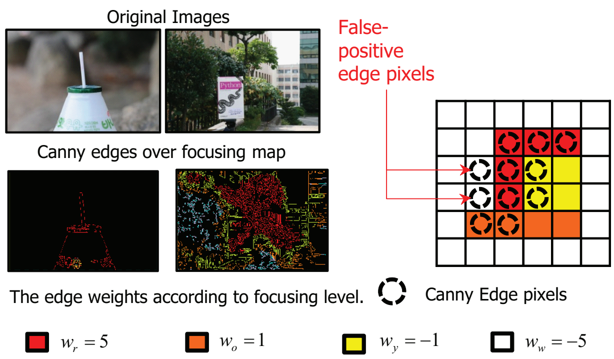

Figure 4: DoF evaluation example. We compute the number of the Canny edge pixels in the mask corresponding to the level of the focusing map. The weight values of the focusing map, w_r, w_o, w_y, and w_w are *5*,*1*,*-1*,*-5*, respectively. The negative weights on the yellow and white regions are false-positive pixels, since they represent the unfocused region.

$$f(p) = \sum_{np \in m} \frac{pixels(Canny(np))}{|dist(p,np)|}, \qquad (1)$$

where $dist(p_i, p_j)$ means the Manhattan distance between two pixels, p_i and p_j. $Canny()$ depicts the pixel sets extracted from Canny edge detection algorithm. Note that we assume the red (R_r) and orange region (R_o) is in-focus, whilst the other regions are out-of-focus. The other regions contain information that is less important than R_r and R_o, since there are less edge pixels. However, we also need the other regions (R_y, R_w) to determine the false-positive case, as shown in Figure 4. Finally, we calculate the total score of the edge pixels according to the level of focus by :

$$E_d(P) = \sum_{p \in P} (w_k(p) \cdot f(p)) / pixels(e_c(P)), \qquad (2)$$

where e_c, $pixels(I)$ and $w_k(p)$ are the extracted edge pixels using Canny edge detection, the number of pixels stored in region I and the weights according to the focusing level, respectively. Figure 5 shows the result of our DoF assessment. You can see that the focused regions (R_r and R_o) are becoming clustered on the central area (near the rabbit objects) from left to right in the sequence .

4 EXPERIMENTS

We invited twelve digital camera users, divided into two groups based on their photo taking ability. The beginner group, $G1$, consists of four camera users who can take pictures using their compact digital camera controlling the embedded camera modes, for example, M (Manual), A (auto Aperture), S (auto Shutter speed) and P (all Programmed) modes. The expert group $G2$ of six users could control the detailed camera parameters, such as aperture, shutter speed, and ISO. They also have sufficient experience, shotting pictures using their DSLR cameras for more than two years.

We collected 206 photos from five categories, as described in Table 1. The general photo sets A and B consist of random shots of photos taken, while in motion (A), and still photos (B). Several experiment conditions were controlled for the experimental photo sets. We used DSLR cameras, controlling camera parameters, such as aperture and shutter speed, to shoot photo set C. The camera was shaken when we took photos for the final photo set, D.

The participants in our user study were asked to construct three photo sets according to the photo quality. The input photo sets used in this experiment were 30

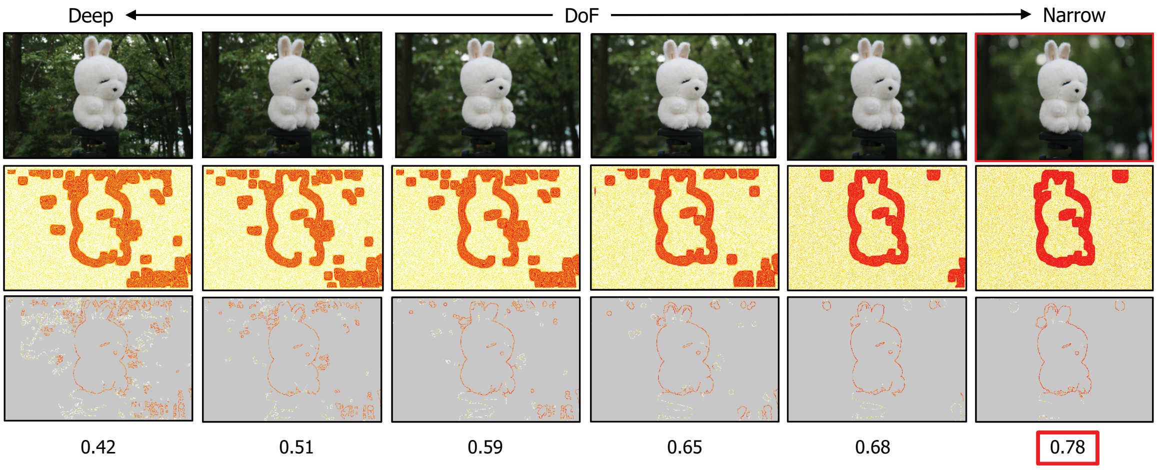

Figure 5: Evaluation results of contrast and DoF features. Experimental photos (*C*, *D*, and *E* in Table 1) for DoF. We took these photos controlling shutter speed and aperture parameters. The focus regions become clustered on the central object from left to right.

Purpose (♯ of photos)	Sets	Taken method	♯ of photos
General photos (100)	A	traveling	69
	B	still shots	31
Experimental photos (106)	C	DoF	60
	D	shaking photos	46

Table 1: Input photo sets used in this experiment.

User	eval. class	precision		recall		F-score
		avg.	std.	avg.	std.	
G1	upper	0.75	0.12	0.88	0.14	0.81
	middle	0.77	0.14	0.82	0.07	0.79
	lower	0.87	0.07	0.87	0.08	0.87
G2	upper	0.86	0.08	0.85	0.09	0.85
	middle	0.85	0.10	0.89	0.07	0.87
	lower	0.84	0.06	0.85	0.05	0.84
avg. score		0.82	0.11	0.87	0.10	0.84

Table 2: Photo quality assessment. The photo sets used are described in Table 1. We investigate precision and recall with their manual operation outcome to evaluate our metrics.

randomly selected from each of the shuffled input photo sets.

We calculate the average precision, and recall, as described in Table 2, to compare our classification result to those classified manually by the users. Precision and recall are two widely used metrics for evaluating the correctness of a pattern recognition algorithm. When using precision and recall, the set of possible labels for a given instance is divided into two subsets, one of which is considered “relevant" for the purposes of the metric. Recall is then computed as the fraction of correct instances among all instances that actually belong to the relevant subset, while precision is the fraction of correct instances among those that the algorithm believes to belong to the relevant subset. A measure that combines precision and recall is the harmonic mean of precision and recall. Since F-score is the measure recall and precision are evenly weighted, it means the reliability of their experiment. The proposed measure efficiently assesses DoF (its precision and recall exceed 0.82). Especially, the expert group agreed that our photo assessment measure is useful in evaluating the professional photos. However, since the beginner group, G1, prefers the high contrast photos, their experiment results (precision and recall) are somewhat poor (both values are less than 0.87).

5 CONCLUSION

In this paper, we proposed a novel method to assess photo quality based on a focusing map. The focussing map is a form of a saliency map that is classified by Canny edge distribution. Its goal is to simulate the attention mechanism of a human visual system. This is simple to implement. It is implemented in CUDA to decrease the time required for image processing for a large resolution. We also conducted an experiment based on the precision, recall and F-score, with four photo sets (206 photos) to compare our performance with a user's manual evaluation. The experiment shows the proposed measure perform well compared to manual user evaluation (The precision and recall exceed 0.8). Especially, the expert group agreed that our photo assessment metrics are useful to evaluate each photo.

In this paper, we considered only one criteria to assess the Depth of Field. However, it is insufficient to assess general photos, since there are many features to be considered (for example, blur, contrast and exposure). Measurements developed for other features would need a combination method for multiple features, for example, Ke's naive Bayes classifier [2]. This method enables various metrics that are dependent on each other. Therefore, we propose the development of further metrics for photo quality evaluation (for example, blur and contrast) and a combination method to integrate these metrics naturally. In this paper, we conducted on naive experiment to find the four classified levels of four focusing region, R_r, ...,R_w. Although, we deduced it from a naive experimental result, it is also important to find the scientific reason of the four classified level ratio.

ACKNOWLEDGMENTS

This work was supported by the IT R&D program of MKE/MCST/KEIT (KI001820, Development of Computational Photography Technologies for Image and Video Contents).

REFERENCES

[1] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Studying aesthetics in photographic images using a computational approach. In *In Proc. of ECCV*, pages 7–13, 2006.

[2] Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 419–426, 2006.

[3] Renting Liu, Zhaorong Li, and Jiaya Jia. Image partial blur detection and classification. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 1 –8, 23-28 2008.

[4] Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In *Proc. of the 10th European Conference on Computer Vision*, pages 386–399, Berlin, Heidelberg, 2008.

[5] Xiaoshuai Sun, Hongxun Yao, Rongrong Ji, and Shaohui Liu. Photo assessment based on computational visual attention model. In *Proc. of the 17th ACM MM*, pages 541–544, 2009.

[6] Niranjan Damera Venkata, Thomas D. Kite, Wilson S. Geisler, Brian L. Evans, and Alan C. Bovik. Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 9(4):636–650, 2000.