Flash Cut: Foreground Extraction with Flash and No-flash Image Pairs

Jian Sun[1] Jian Sun[2] Sing Bing Kang[3] Zong-Ben Xu[1] Xiaoou Tang[2] Heung-Yeung Shum[2]

[1]Xi'an Jiaotong University
Xi'an, P. R. China

[2]Microsoft Research Asia
Beijing, P. R. China

[3]Microsoft Research
Redmond, WA, USA

Abstract In this paper, we propose a novel approach for foreground layer extraction using flash/no-flash image pairs, which we call *flash cut*. Flash cut is based on the simple observation that only the foreground is significantly brightened by the flash and the background appearance change is very small, if the background is distant. Changes due to flash, motion, and color information are fused in an MRF framework to produce high quality segmentation results. Flash cut handles some amount of camera shake, and foreground motion, which makes it practical for anyone with a flash-equipped camera to use. We validate our approach on a variety of indoor and outdoor examples.

1. Introduction

Image segmentation is a fundamental problem in computer vision. However, achieving a good or perfect segmentation result from a single image, even for foreground/background separation, is still challenging. There are two basic types of approaches for segmentation. One is interactive image segmentation [7, 14, 21] with minimal user assistance. The other relies on the use of additional information and/or more images, such as motion [4, 27, 31, 32], stereo [9], and infrared light [6], or the assumption of a known and static background [26]. In general, approaches that use multiple images tend to produce better and more robust results [2, 8]. To be practical, however, we want to minimize the number of shots necessary (e.g., two shots) and be able to handle the inevitable appearance changes due to camera motion (caused by hand shake) and scene motion.

In this paper, we propose a practical approach to automatically extract high-quality foreground layer using a flash/no-flash image pair. Figure 1(a) shows a typical portrait/travel photo—a near subject against a distant background scene. Automatic foreground extraction from a single image is non-trivial because of the arbitrary color distributions of the foreground and background. Figure 1(b) is another photo taken with flash immediately after the first one. There are misalignments between the two photos because of small camera and subject motion. Notice that only the foreground is significantly brightened by the flash and the background appearance change is very small. This is due to the rapid flash intensity falloff with distance. We use this simple observation to design a technique, which we call *flash cut*, for producing high-quality foreground extraction results. Figure 1(c) and (d) show the new composition results generated by our technique.

Active lighting has long been exploited for segmentation. In the film industry, sodium or ultraviolet light [29, 12] is used in a well-controlled studio environment and recorded on an additional strip of film. Image or video matting relies on this extra film. The "SegCam" system [6], by turning on and off an infrared LED, tags brightened pixels as foreground. The flash/no-flash idea has been used for non-photorealistic rendering [20]; here, depth edges are detected through shadows cast by spatially distributed flashes. This idea is extended for stereo vision [11] by requiring shadows be cast and observed on the background nearby. The work most related to ours is flash matting [25], in which a flash/no-flash image pair is used to automatically produce very accurate matting results. However, flash matting generally requires the two input images to be pixel aligned and the scene to be static. Techniques that capitalize on the flash tend to be easy to implement (and thus practical), as the typical off-the-shelf camera is equipped with flash.

A common drawback of approaches that rely on active lighting is that images taken under different lighting conditions must be well-aligned and/or image capture is limited to indoor scenes. In our work, we combine flash, motion, and color cues in an MRF framework to handle moderate amounts of camera and subject motion. Our technique makes it more practical to obtain high-quality segmentation results for both indoor and outdoor scenes.

1.1. Related work

There are many approaches that use motion information to separate out the foreground [3, 4, 27, 30, 31, 32]. Such approaches typically require motion to be segmented or grouped. Unfortunately, foreground and background motions are similar for small parallax. On the other hand, if the parallax is large, there is the difficult issue of han-

(a) no-flash image (b) flash image (c) composition after flash cut (d) composition after flash cut

Figure 1. Images taken sequentially with the camera flash off (a) and then on (b). (c,d) are the results of applying flash cut and pasting the extracted foregrounds onto new backgrounds.

dling occlusions and disocclusions. Textureless regions and non-rigid motion also complicate the segmentation problem. Because motion-based techniques rely on neighborhood information for grouping, foregrounds with thin structures are particularly difficult to extract. Examples of such foregrounds can be seen in Figures 6 and 7.

Also related to our work is background substraction or modeling. To relax the requirement of a known and static background, many approaches [13, 17, 18, 23, 28] have been proposed to handle background changes or dynamic backgrounds. Usually, these approaches require an image sequence (e.g., 8-15 frames) so that the background pixels are visible in parts of the frames. The results produced by these approaches are typically not accurate enough for high-quality foreground extraction.

Flash-based techniques have also been used in computer graphics. Examples include enhancement of flash photos [10] and removal of artifacts and flash-exposure sampling [1]. Other applications that rely on flash/no-flash information include denoising, detail transfer, white balancing, continuous flash, and red-eye correction [19].

1.2. Our approach

Our flash cut technique handles the segmentation problem by fusing flash, motion, and color cues. The flash information is exploited globally using histogram analysis and locally with the help of background motion estimation. The segmentation problem is formulated in an MRF framework which can be efficiently solved by min-cut. Our approach is insensitive to misalignments caused by camera shake and small foreground object movement. It is also able to handle highly complex objects such as plants and trees. The biggest advantage of our approach lies in its simplicity—it is easy to automatically extract high-quality foreground objects using an off-the-shelf, flash-equipped camera.

2. Problem Formulation

In this section, we describe our flash cut technique, which capitalizes on the flash imaging model.

2.1. Flash imaging model

Assuming the flash is a point light source with intensity L, the radiance E of a surface point P caused by the flash is $E = L \cdot \rho \cdot r^{-2} \cdot \cos\theta$, where ρ is the surface BRDF under given flash and view directions, r is distance from the flash unit, and θ is the angle between flash direction and surface normal at P. Hence, the flash intensity falls off quickly with distance r.

In this paper, we assume that the background layer is distant compared with the foreground layer and camera flash. Under this assumption, the appearance of the foreground will be dramatically changed by the flash while the background appearance is marginally changed. Figure 1(a) and (b) show a no-flash/flash image pair; this pair is a typical example under our assumption. The change difference between the foreground and the background caused by the flash provides us a very strong cue for the foreground layer extraction.

2.2. Segmentation model

Foreground/background segmentation can be formulated as a binary labeling problem. Given one of the two input images, i.e., flash image I^f or no-flash image I^{nf}, the goal is to label pixel p by $x_p \in \{\text{background}(=0), \text{foreground}(=1)\}$. The foreground layer is extracted by minimizing the following energy of an Markov Random Field (MRF):

$$E(X) = \sum_p E_d(x_p) + \alpha \sum_{p,q} E_s(x_p, x_q), \qquad (1)$$

where $E_d(x_p)$ is the data term for each pixel p, and $E_s(x_p, x_q)$ is the smoothness term associated with two adjacent pixels p and q. The parameter α balances the influence of these two terms. In our experiments, the values used range between 20 and 40.

The smoothness $E_s(x_p, x_q)$ penalizes the different labeling for two adjacent pixels p and q in textureless areas. It is defined as

$$E_s(x_p, x_q) = |x_p - x_q| \cdot \exp(-\beta ||I_p - I_q||^2), \qquad (2)$$

where $\beta = (2\langle ||I_p - I_q||^2 \rangle)^{-1}$ [5] and $\langle \cdot \rangle$ indicates expectation. The energy (1) with this kind of contrast dependent smoothness term is originally formulated and solved using graphs cut by [7].

The data term $E_d(x_p)$ models the flash effects on the foreground, the (motion compensated) background, and the color likelihood. It consists of three terms:

$$E_d(x_p) = \gamma_f E_f(x_p) + \gamma_m E_m(x_p) + E_c(x_p). \qquad (3)$$

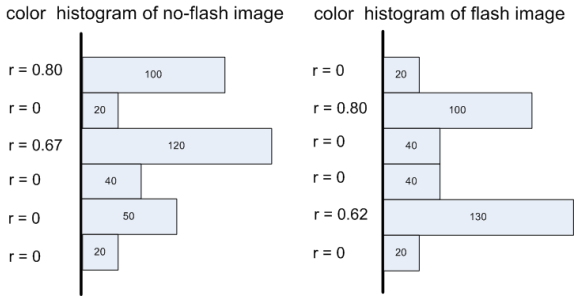

Figure 2. Illustration of flash ratio for a flash/no-flash pair. For each histogram, the intensity is increasing downward, with the number of pixels indicated within each bin. The left of each bin is the flash ratio. In the top-most bin, there are 100 pixels in the no-flash image and 20 in the flash image. Hence, the flash ratio = (100-20)/100 = 0.80. In this case, 80 pixels are brightened by the flash and moved out onto the other bins in the flash image. For the fifth bin (from top), there is a 62% addition to the number of pixels in the flash image caused by pixels brightened by the flash.

γ_f and γ_m are both set to 10 in all our experiments.

- E_f is the foreground flash term, which tends to label the pixel with significant appearance change as foreground. This energy term uses the color histogram of two images as a global cue (Section 3.1).
- E_m is the background flash term, which models the motion-compensated flash effect on the background. It tends to label the pixel with good matching and small appearance changes as background. This energy term considers both flash and motion cues (Section 3.2).
- E_c is the color term, which models the foreground and background color likelihoods in the image. The foreground and background color distributions are modeled as Gaussian Mixture Models (Section 3.3).

3. Flash Cut

In this section, we describe the details of three energy terms defined in the previous section.

3.1. Foreground flash term

We model the flash effect on the foreground by analyzing the histograms of the flash/no-flash images. This is global information on changes caused by flash, which is insensitive to small camera and scene movements.

Let $H^f = \{h_k^f\}$ and $H^{nf} = \{h_k^{nf}\}$ be the RGB color histograms of the flash image and the no-flash image, respectively; h_k^f and h_k^{nf} are their respective pixel counts in the kth bin. If $h_k^{nf} > h_k^f$, some pixels in the kth bin of H^{nf} are sufficiently modified by the flash and moved to other bins in H^f (these bins are unknown). As a result, the pixels in this bin for the no-flash image has a higher probability to be foreground pixels. Similarly, $h_k^{nf} < h_k^f$ means that some pixels have been modified by the flash and transferred to the kth bin of H^f. Hence, in the flash image, the pixels in this bin have a higher probability to be foreground pixels. We quantify these simple observations by defining the *flash ratio* for each pixel p in the flash/no-flash images as

$$r_p^f = \max\{\frac{h_{k_p}^f - h_{k_p}^{nf}}{h_{k_p}^f}, 0\}, \quad r_p^{nf} = \max\{\frac{h_{k_p}^{nf} - h_{k_p}^f}{h_{k_p}^{nf}}, 0\},$$

where k_p is the bin index of the pixel p. The larger the flash ratio for a pixel is, the higher probability the pixel belongs to the foreground. Figure 2 is a 1D illustration of flash ratio.

Figure 3(b) shows the flash ratios of the no-flash/flash image pair. Because the histogram is global, the flash ratio map may not be entirely correct. For example, the flash ratio of the ground in the no-flash image is high because the color of the ground is similar to the color of clothes in the no-flash image. The flash ratios of the black eye-glasses in both images are low due to the low reflectance of the black object.

We define the energy term based on the flash ratio with a robust parameter ζ as

$$E_f(x_p) = \begin{cases} 0 & , x_p = 1 \\ \frac{1}{1-\zeta}[\max\{r_p, \zeta\} - \zeta] & , x_p = 0 \end{cases}. \quad (4)$$

The default value of ζ is set to 0.2. The significance of this robust parameter is the following: if r_p is larger than ζ, we are more likely to label pixel p as the foreground. Otherwise, the costs for labeling pixel p as the foreground and background are the same. Thus, the energy term $E_f(x_p)$ provides a conservative estimate of the foreground layer.

3.2. Motion-compensated background flash term

Suppose we have dense motion field $\mathbf{m} = \{m(p)\}$ that registers the no-flash image I^{nf} to the flash image I^f. The flash difference between the pixel p in I^{nf} and its corresponding pixel $p' = m(p)$ in I^f is

$$\triangle I_p = I_{m(p)}^f - I_p^{nf} = I_{p'}^f - I_p^{nf}. \quad (5)$$

Since the user is expected to capture the flash/no-flash images with distant background in quick succession, the appearance change of the background is expected to be small and uniform. It is thus reasonable to model the flash difference distribution of all background pixels as a Gaussian distribution $N(\triangle I_p|\mu, \sigma^2)$ with mean μ and variance σ^2. Then, we can define the probability of a pixel p belonging to the background as

$$p_b(x) = \exp(-\sigma_b(\triangle I_p - \mu)^2). \quad (6)$$

We set $\sigma_b = \ln 2/(3\sigma)^2$ so that the pixel with flash difference within the $\pm 3\sigma$ interval around μ will be given a higher probability (≥ 0.5) to be background. We call $p_b(x)$

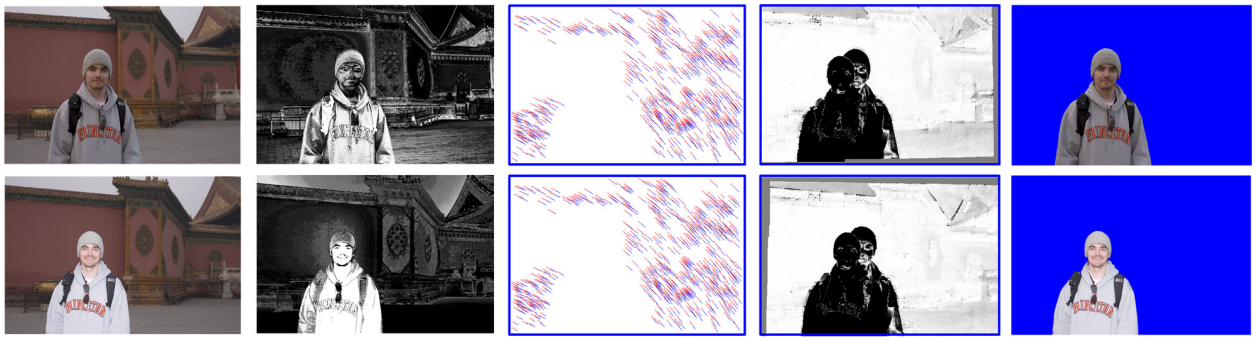

Figure 3. Intermediate results of the flash/no-flash image pair in Figure 1. (a) input images. (b) flash ratio maps. (c) one-to-one sparse matching (top) and background sparse matching (bottom). (d) background probability maps. (e) results by flash cut. In (d), the brighter the pixel, the higher the probability.

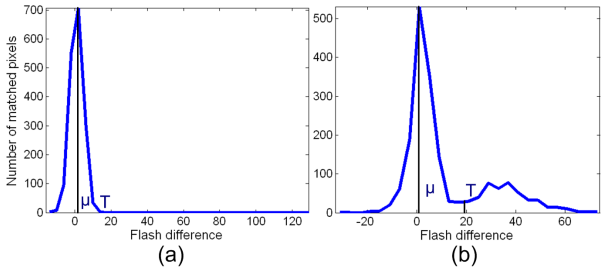

Figure 4. Flash difference histograms of sparse matched features for two different flash/no-flash pairs. (a) Histogram with one peak, and (b) histogram with multiple peaks. The parameter μ is estimated using the first maximum, and T is the first minimum on the right of μ.

the background probability, as shown in Figure 3(d). Finally, the energy term $E_m(x_p)$ is defined as

$$E_m(x_p) = \begin{cases} 2\max\{p_b(x_p), 0.5\} - 1 & , x_p = 1 \\ 0 & , x_p = 0 \end{cases}. \quad (7)$$

With this definition, $E_m(x_p)$ is normalized to be in the range $[0, 1]$. The energy for the flash image is similarly defined.

In the above definition, we assume that we know the parameters $\{\mu, \sigma^2\}$ and the dense motion field $\mathbf{m}$. We now describe how these parameters are estimated.

Parameter estimation. We estimate the parameters $\{\mu, \sigma^2\}$ by analyzing the one-to-one sparse feature matching between two images. The one-to-one sparse feature matching is established using the SIFT detector [15] and nearest-neighbor matching. We require the nearest neighbor to be significantly better than the second-nearest neighbor (we set the distance ratio threshold as 0.6). The uniqueness of matching is enforced using cross-checking. Figure 3(c) shows the matched sparse features for the image pair.

Given matched sparse features, we construct the 1D histogram of the flash difference. In most cases, only a few or even no matched features are from the foreground layer because of the dramatic appearance change of the foreground, as shown in Figure 3(c). The corresponding flash difference histogram is shown in Figure 4(a). In some cases, the matched features come from both the foreground and background, resulting in multiple peaks in the histogram. For example, the histogram in Figure 4(b) is constructed from the matching between the image pair in Figure 6. In both cases, we are only concerned about the first local peak, which corresponds to background matches. The flash difference mean μ is estimated by the first local maximum bin in the histogram. The flash difference variance σ^2 is estimated using the matched features whose flash difference is lower than a threshold T. We set the threshold T at the first local minimum bin above μ, as shown in Figure 4. The histogram is smoothed using a Gaussian kernel to reduce noise.

Motion estimation. Because our purpose is to only estimate the background motion, we first discard matched sparse features with flash difference above the threshold T (the same parameter in the previous paragraph). For example, the bottom of Figures 3(c) and 5(d) show the background sparse matching. For the case where the background is stationary and distant, e.g., Figure 7(a) and (c), a global motion model such as a homography or affine is sufficient to model the background motion. We can directly compute a global motion from the matched background sparse features. However, a global motion is inadequate to model a dynamic background, radially distorted images, or parallax, e.g., Figure 5 with a mildly dynamic background. In these cases, applying dense motion refinement will improve the result.

Taking the no-flash image as an example, we refine the initial dense background motion field $\mathbf{m}^0$ interpolated by Adaptively Locally Weighted Regression (ALWR) [15] using matched background sparse features. The matching residual with compensated flash difference is

$$e_p = I^f_{m(p)} - I^{nf}_p - \mu_p. \quad (8)$$

Recall that I^f and I^{nf} refer to the flash image and no-flash image, respectively.

For each pixel p, its initial flash difference μ_p is set as μ. Then, the motion field is iteratively refined using the Lucas-Kanade [16] algorithm. The motion correction $\triangle m^k(p)$ in iteration k is estimated by

$$\triangle m^k(p) = -(\sum_{q\in w(p)} \nabla I^f_{m^k(q)} \nabla I^{f^T}_{m^k(q)})^{-1} \sum_{q\in w(p)} \nabla I^f_{m^k(q)} e^k_q.$$

$w(p)$ is the 5×5 window around the pixel p. After convergence, we re-estimate μ_p for each pixel locally in a 11×11 window $w'(p)$ by $\mu_p = \frac{1}{11\times 11}\sum_{q\in w'(p)}(I^f_{m(q)} - I^{nf}_q)$. Then, the Lucas-Kanade algorithm is run again to further refine the motion field. The iteration number is 2 or 3 in our experiments.

Figure 5(e) and (f) show the background probability map of the flash image by a global homography, and our refined dense motion field. (The brighter the pixel, the higher the background probability.) Clearly, the background probability map is improved with the use of the motion field. Notice that while the foreground layer may be incorrectly registered, it will always be assigned a lower probability due to the significant foreground luminance changes.

3.3. Color term

The foreground color likelihood is modeled as Gaussian Mixture Models (GMMs) [5]:

$$p_c(I_p|x_p=1) = \sum_{k=1}^{K} w^f_k \mathcal{N}(I_p|\mu^f_k, \Sigma^f_k), \qquad (9)$$

where $N(\cdot)$ is a Gaussian distribution and $\{w^f_k, \mu^f_k, \Sigma^f_k\}$ represent the weight, mean, and covariance matrix of the kth component of the foreground GMMs. The typical values of component number K is 10 in our implementation. The background GMMs is estimated using all pixels with $p_b(x) > 0.6$. The foreground color likelihood $p_c(I_p|x_p=0)$ is similarly defined and estimated using all pixels with $p_b(x) < 0.4$. Finally, the color term $E_c(x_p)$ is defined as

$$E_c(x_p) = \begin{cases} -\log(p_c(I_p|x_p=1)) & , x_p = 1 \\ -\log(p_c(I_p|x_p=0)) & , x_p = 0 \end{cases}. \qquad (10)$$

The color models can be refined after minimizing the flash cut energy, using the newly estimated foreground/background area. However, we found that the improvement is marginal because our initial color models are generally accurate enough.

4. Experimental Results

Figure 6 shows the intermediate results of flash cut and comparison with several other approaches. The top two rows of Figure 6 are input image, flash ratio map, motion compensated background probability map, and flash cut result for the flash/no-flash pair. To see the contributions of the foreground flash term and background flash term, Figure 6 (e) and (f) show unsatisfactory segmentation results using (flash ratio + color) terms and (background prob. + color) terms. It demonstrates the necessity of combining all cues.

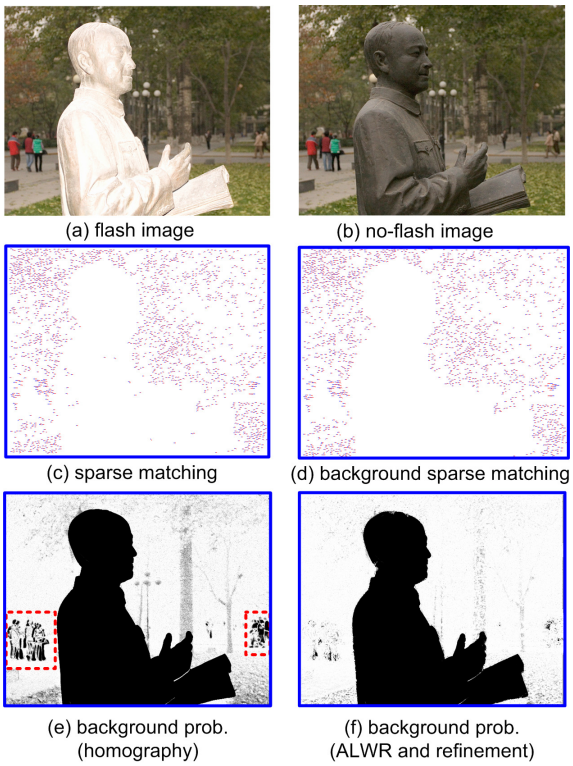

Figure 5. Motion-compensated background probability. (a) Flash image, (b) no-flash image, (c) sparse matching, (d) background sparse matching, i.e., matching with flash difference lower than threshold T. (e) Background probability map for the flash image by a global homography. Areas highlighted within red (dotted) rectangles are areas incorrectly matched due to the background being dynamic. (f) Background probability map by ALWR and refinement.

We also compared flash cut with two other state-of-the-art segmentation algorithms, namely, GrabCut [21] and co-segmentation [22]. Figure 6 (g) shows the segmentation results by GrabCut applied to each image individually, which is not accurate due to color ambiguity. We used the whole image boundary as the initial rectangle as the input to GrabCut. Co-segmentation is a histogram-based algorithm to segment the regions with the same color distribution from an image pair. In our implementation of co-segmentation, we used the ground-truth background color distribution of one image to infer the segmentation in another image. The results in Figure 6 (h) demonstrate the inadequacy of using the global histogram to segment thin structures.

We have tested our approach on a variety of indoor and outdoor flash/no-flash image pairs. Figure 7(a) shows an example of a walking person. Motion-based segmentation

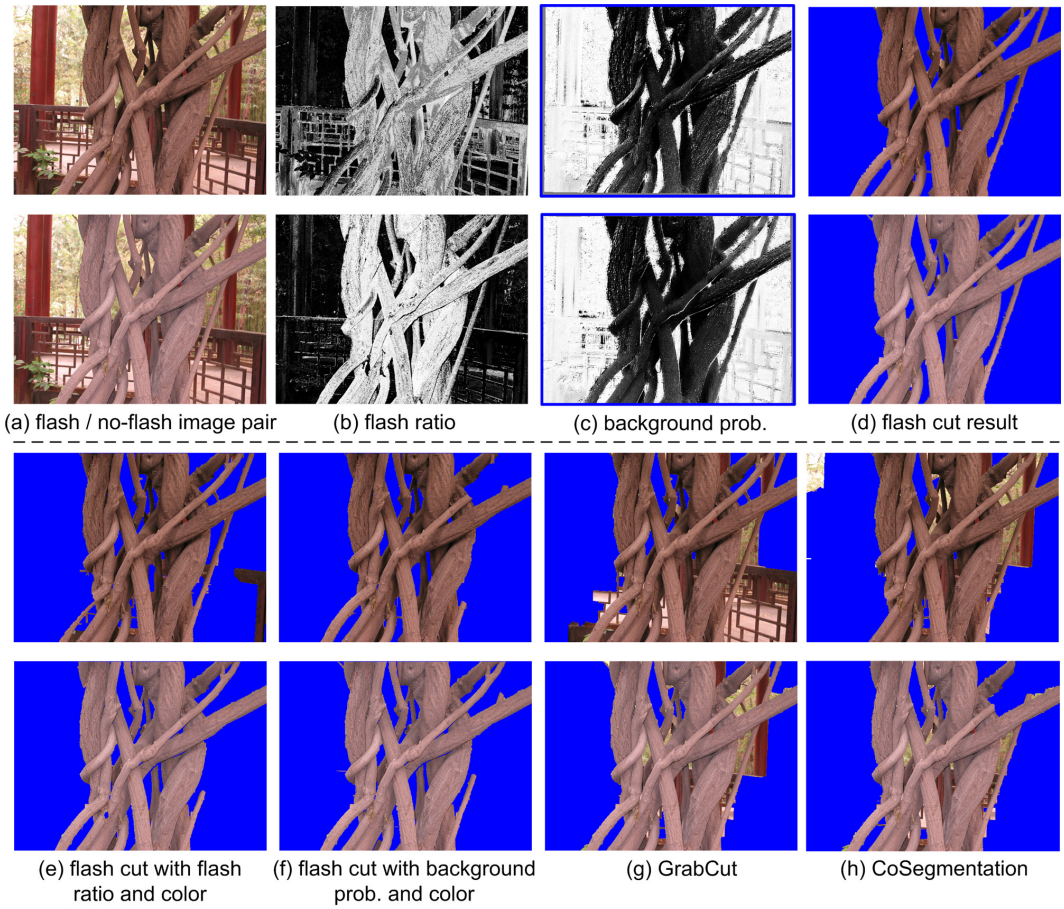

Figure 6. Comparison of results. (a-d) Input images, flash ratio maps, background probability maps, and final flash cut results, (e) flash cut with flash ratio and color only, (f) flash cut with background probability and color only, (g) GrabCut, and (h) CoSegmentation. By combining flash ratio, background probability, and color, our flash cut technique yielded the most accurate result.

techniques would find this example very challenging because of the large motion. Figure 7(b) shows an image pair with a moving background, while Figure 7(c) demonstrates an example of a foreground with fine structures. It is hard for GrabCut, co-segmentation, and motion-based algorithms to achieve satisfactory results for all of these examples simultaneously. Our flash cut was able to extract the detailed structures well. In Figure 7(d), the foreground is a plant (with complicated structures) captured under windy conditions. Figure 7(e) is the flash cut result for the scene with a fence in a bamboo grove; here, the background is complicated. Figure 7(f) is an indoor image pair with both foreground movement and mildly dynamic background. Due to the limited space of the indoor environment, it is necessary to decrease the flash intensity to only light the foreground objects. Figure 7(g) shows another pair of segmentation results in an indoor environment. In these examples, we used the coherent matting [24] on the trimap computed by eroding and dilating flash cut result to produce a more natural composition result.

5. Conclusion

In this paper, we proposed a foreground layer extraction approach using flash/no-flash image pair. By combining flash, motion, and color information, flash cut is able to produce high-quality segmentation results and tolerate small camera and scene motions. It is thus practical and easy for the average user with a flash-equipped camera to use.

Since we formulated segmentation as a binary labeling problem, our approach cannot handle the foreground object with furry, long hair, or transparent boundaries. Combining our technique with a sophisticated matting technique is a possible solution. Like other flash-based techniques, our approach is sensitive to large amounts of self-shadowing in the foreground (small shadow regions can be handled by the smoothness term, e.g., thin shadow regions in Figures 3, 5, 6, and 7 (d) (f)). The flash constraint can be made stronger since the flash chroma is typically the same. In other words, the pixels are not arbitrarily brightened, but rather, brightened with a specific chroma contribution (with different magnitudes). This assumes, of course, that the camera radiometric response is approximately linear and

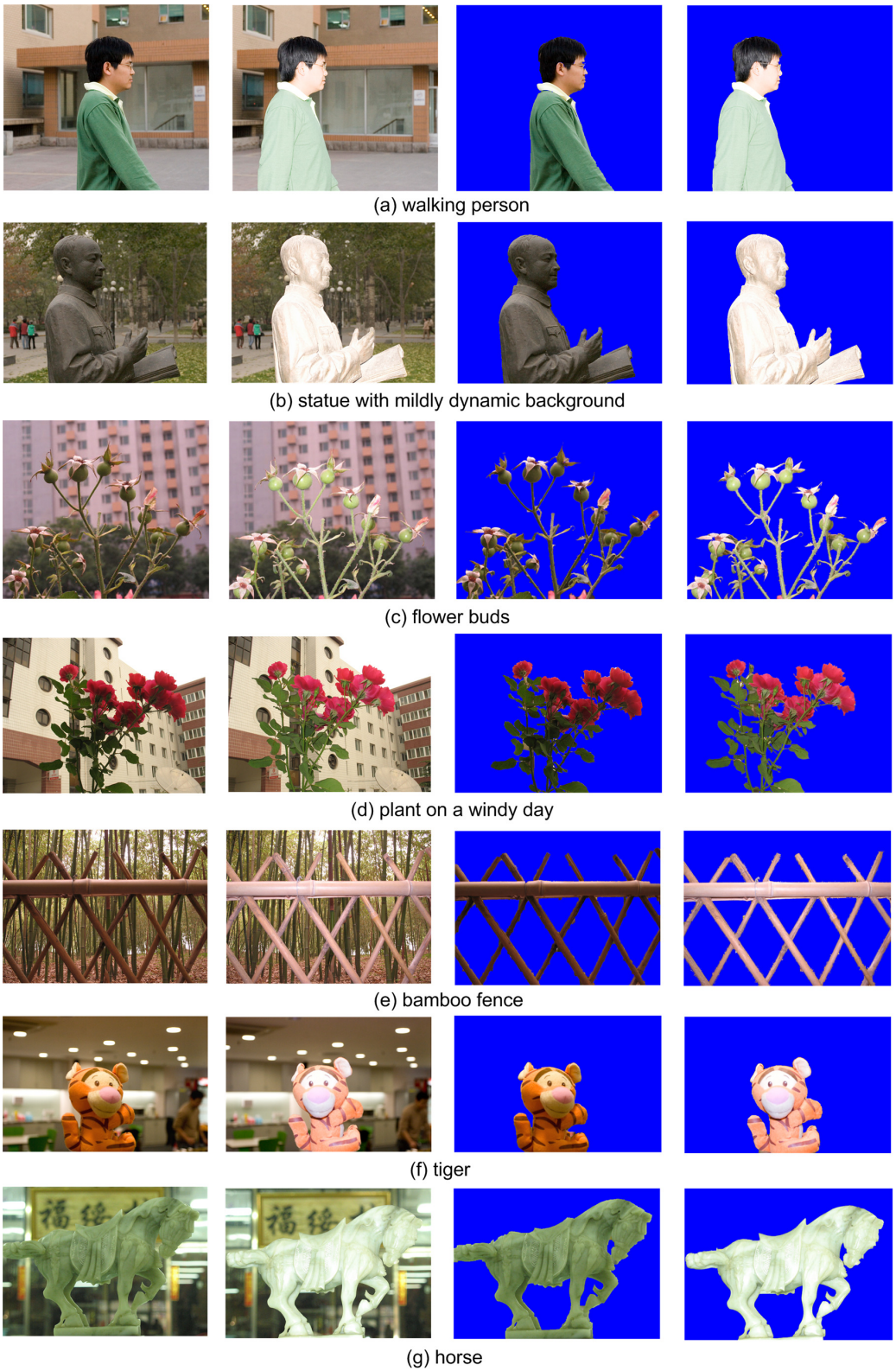

Figure 7. Flash cut results. From left to right: no-flash image, flash image, segmentation results on no-flash image and flash image.

there's no saturation. Another future direction is to extend the flash cut idea to video segmentation.

Acknowledgments. We thank the anonymous reviewers and area chairs for helping us to improve this paper. This work is performed when the first author visited Microsoft Research Asia. The first author and Zong-Ben Xu were supported by a grant from the National Science Natural Foundation of China (No.70531030).

References

[1] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Z. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In *Proceedings of SIGGRAPH*, pages 828–835, 2005.

[2] N. Apostoloff and A. Fitzgibbon. Bayesian video matting using learnt image priors. In *Proceedings of CVPR*, volume I, pages 407–414, 2004.

[3] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. In *IEEE Trans. on PAMI*, volume 14, pages 886–896, 1992.

[4] P. Bhat, K. C. Zheng, N. Snavely, A. Agarwala, M. Agrawala, M. F. Cohen, and B. Curless. Piecewise image registration in the presence of multiple large motions. In *Proceedings of CVPR*, volume II, pages 2491–2497, 2006.

[5] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr. Interactive image segmentation using an adaptive GMMRF model. In *Proceedings of ECCV*, volume I, pages 428–441, 2004.

[6] R. M. Bolle, J. H. Connell, N. Haas, R. Mohan, and G. Taubin. Object imaging system. *U. S. Patent 5,631,976*, 1994.

[7] Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In *Proceedings of ICCV*, 2001.

[8] Y. Y. Chuang, A. Agarwala, B. Curless, D. H. Salesin, and R. Szeliski. Video matting of complex scenes. In *Proceedings of ACM SIGGRAPH 2002*, pages 243–248, 2002.

[9] A. Criminisi, G. Gross, A. Blake, and V. Kolmogorov. Bilayer segmentation of binocular stereo video. In *Proceedings of CVPR*, volume I, pages 53–60, 2006.

[10] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. In *Proceedings of SIGGRAPH*, pages 673–678, 2004.

[11] R. Feris, R. Raskar, L. B. Chen, K. Tan, and M. Turk. Discontinuity preserving stereo with small baseline multi-flash illumination. In *Proceedings of ICCV*, volume I, pages 412–419, 2005.

[12] R. Fielding. The technique of special effects cinematography. In *Focal/Hastings House, London, 3rd edition*, pages 220–243, 1972.

[13] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *Proceedings of CVPR*, pages 22–29, 1998.

[14] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum. Lazy snapping. In *Proceedings of SIGGRAPH*, 2004.

[15] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[16] B. Lucas and T. Kanade. An iterative image registration technique with application to stereo vision. In *Proc. Intl. Joint Conf. Artificial Intelligence*, pages 674–679, 1981.

[17] A. Mittal and N. Paragios. Motion-based background subtraction using adaptive kernel density estimation. In *Proceedings of CVPR*, pages 302–309, 2004.

[18] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh. Background modeling and subtraction of dynamic scenes. In *Proceedings of ICCV*, pages 1305–1312, 2005.

[19] G. Petschnigg, M. Agrawala, H. Hoppe, R. Szeliski, M. Cohen, and K. Toyama. Digital photography with flash and no-flash image pairs. In *Proceedings of SIGGRAPH*, pages 664–672, 2004.

[20] R. Raskar, K. Tan, R. Feris, J. Y. Yu, and M. Turk. Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging. In *Proceedings of SIGGRAPH*, pages 673–678, 2004.

[21] C. Rother, A. Blake, and V. Kolmogorov. GrabCut - Interactive foreground extraction using iterated graph cuts. In *Proceedings of SIGGRAPH*, 2004.

[22] C. Rother, V. Kolmogorov, T. Minka, and A. Blake. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into MRFs. In *Proceedings of CVPR*, volume I, pages 993–1000, 2006.

[23] Y. Sheikh and M. Shah. Bayesian object detection in dynamic scenes. In *Proceedings of CVPR*, pages 1778–1792, 2005.

[24] H. Y. Shum, J. Sun, S. Yamazaki, Y. Li, and C. K. Tang. Pop-up light field: An interactive image-based modeling and rendering system. *ACM Transaction of Graphics*, 23(2):143–162, 2004.

[25] J. Sun, Y. Li, S. B. Kang, and H. Y. Shum. Flash matting. In *Proceedings of SIGGRAPH*, pages 361–366, 2006.

[26] J. Sun, W. Zhang, X. Tang, and H. Y. Shum. Background cut. In *Proceedings of ECCV*, volume II, pages 628–641, 2006.

[27] P. H. S. Torr, R. Szeliski, and P. Anandan. An integrated Bayesian approach to layer extraction from image sequences. *IEEE Trans. on PAMI.*, 23(3):297–303, 2001.

[28] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers. Wallflower: principles and practice of background maintenance. In *Proceedings of ICCV*, pages 255–261, 1999.

[29] P. Vlahos. Composite photography utilizing sodium vapor illumination. *U. S. Patent 3,095,304*, 1958.

[30] J. Y. A. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proceedings of CVPR*, pages 361–366, 1993.

[31] J. Wills, S. Agarwal, and S. Belongie. What went where. In *Proceedings of CVPR*, volume I, pages 37–44, 2003.

[32] J. J. Xiao and M. Shah. Motion layer extraction in the presence of occlusion using graph cut. In *Proceedings of CVPR*, volume II, pages 972–979, 2004.