Flash Matting

Jian Sun[1] Yin Li[1] Sing Bing Kang[2] Heung-Yeung Shum[1]

[1]Microsoft Research Asia, Beijing [2]Microsoft Research, Redmond, WA

Figure 1: Flash matting results for flower scene with complex foreground and background color distributions. From left to right: flash image, no-flash image, recovered alpha matte, and flower basket with new background.

Abstract

In this paper, we propose a novel approach to extract mattes using a pair of flash/no-flash images. Our approach, which we call *flash matting*, was inspired by the simple observation that the most noticeable difference between the flash and no-flash images is the foreground object if the background scene is sufficiently distant. We apply a new matting algorithm called *joint Bayesian flash matting* to robustly recover the matte from flash/no-flash images, even for scenes in which the foreground and the background are similar or the background is complex. Experimental results involving a variety of complex indoors and outdoors scenes show that it is easy to extract high-quality mattes using an off-the-shelf, flash-equipped camera. We also describe extensions to flash matting for handling more general scenes.

1 Introduction

Given a background image B and a foreground image F with its alpha channel or matte α, a new image I can be composed using the *compositing* equation:

$$I = \alpha F + (1 - \alpha)B. \quad (1)$$

Conversely, the goal of image matting is to estimate α, F, and B from a given image I. Despite 40 years of research, making image matting robust and practical remains a continuing challenge.

Image matting from a single image is fundamentally underdetermined. The most common approach for pulling a matte is blue screen matting [Smith and Blinn 1996], in which a foreground object is captured in front of a known solid-colored background, usually blue or green. Blue screen matting is the standard technique employed in the movie and TV industries because a known background B greatly simplifies the matting problem. However, blue screen matting requires an expensive well-controlled studio environment to reduce artifacts such as blue spill, backing shadows, and backing impurities [Smith and Blinn 1996]. In addition, blue screen matting is less suitable for outdoor scenes.

Natural image matting approaches [Berman et al. 2000; Ruzon and Tomasi 2000; Hillman et al. 2001; Chuang et al. 2001; Sun et al. 2004; Wang and Cohen 2005] were proposed to directly compute the matte from a single natural image. First, the input image is manually partitioned into definitely foreground, definitely background, and unknown regions. These three regions are collectively referred to as the *trimap* [Chuang et al. 2001]. Then, α, F, and B are estimated for all pixels in the unknown region. Natural image matting uses additional information from the trimap's spatially coherent distribution of pixels in each region. Unfortunately, they fail to work when the foreground and the background are alike or highly textured. In addition, specifying a good trimap can be labor intensive, especially for complex objects.

In this paper, we propose flash matting for natural image matting using a flash/no-flash image pair. Using additional information from a flash image, our flash matting can extract a high-quality matte even when the foreground and the background have similar colors or when the background is complex. In addition, no special studio environment is required in our approach. Such an example is shown in Figure 1 where both the trimap and matting results were automatically generated.

In flash photography, the flash intensity falls off as inverse distance squared. This is a drawback in flash photography because a distant background may be poorly lit by the flash compared to the foreground (producing the "tunnel" effect). We use the "tunnel" effect for effective image matting. The difference between the flash and no-flash images (we call flash-only image) allows reliable foreground-background separation.

The key contribution of this paper is the application of flash photography to image matting. We made two basic assumptions: 1) only the appearance of the foreground is dramatically changed by the flash; 2) the input image pair is pixel aligned. We developed the joint Bayesian flash matting algorithm to selectively fuse information from two images, e.g.,

flash-only image and no-flash image. To further relax the first assumption, we propose *foreground multi-flash matting* and *background flash matting* to make flash matting more applicable. The second assumption limits our technique to the static scene. This limitation is also shared by many image-based rendering techniques, HDR capture methods, and flash/no-flash photography. In this paper, we proposed *quick flash matting* to allow the capturing for dynamic background.

2 Related Work

Natural image matting from a single image. Most natural image matting approaches are statistical sampling based [Berman et al. 2000; Ruzon and Tomasi 2000; Hillman et al. 2001; Chuang et al. 2001; Wang and Cohen 2005], except for Poisson matting [Sun et al. 2004], which is gradient domain based. Assuming that the image is spatially coherent, statistical sampling based methods use nearby foreground and background colors to estimate α, F, and B for unknown pixels. Gradient domain based method reconstructs the matte from estimated matte gradient by solving Poisson equations. Although local Poisson matting [Sun et al. 2004] is capable of producing state-of-the-art results, it requires a lot of user interaction. To date, given a reasonably good trimap, Bayesian matting [Chuang et al. 2001] is still regarded as one of the most robust methods for automatically generating mattes.

All previous methods that use a single image share three major weaknesses. First, it is hard to resolve ambiguity when the foreground is similar to the background. Second, it is difficult to handle complex foregrounds and backgrounds. Finally, the trimap has to be manually and often carefully specified. Despite these weaknesses, they are the only ones applicable to general images.

Matting from multiple images or video. By taking two images the subject with different backgrounds and two clean background images after removing the subject [Smith and Blinn 1996], image matting is solved in an over-determined system. In video matting [Chuang et al. 2002], after propagating the user-supplied trimaps from the key frames to all the other frames, Bayesian matting is independently applied on each frame.

Defocus video matting [McGuire et al. 2005] is the first fully automatic video matting system for dynamic scenes. Here, video is captured using a special imaging device, which consists of three synchronized cameras that share a virtual optical center but differ in their planes of focus. The trimap is automatically computed by measuring the relative foci; the matte is then solved through constrained optimization. Unfortunately, defocus video matting does not handle foregrounds or backgrounds that are similar or with low frequency textures. The output quality is also sensitive to camera misalignment and color calibration errors.

Active lighting matting. Sodium, infrared, and ultraviolet processes [Vlahos 1958; Fielding 1972] are also called *multi-film matting*. These techniques simultaneously recodes the image on one strip of film and the matte on another strip of film. Perhaps the work most closely related to ours is "SegCam" [Bolle et al. 1994]. Their device works by turning an invisible infrared LED on and off, and then looking for brighter pixels. However, their system is only capable of very rough binary foreground/background separation. In our work, we alter the the foreground appearance using the camera flash, which is considerably more convenient.

Flash photography. Flash-based techniques have been widely adopted in the graphics community in recent years. These include flash photography enhancement [Eisemann and Durand 2004], removal of photography artifacts and flash-exposure sampling [Agrawal et al. 2005], and flash photography applications such as denoising, detail transfer, white balancing, continuous flash, and red-eye correction [Petschnigg et al. 2004]. The goal of these techniques is to generate a new high-quality image from flash/no-flash images. In multiple flash imaging, the shadow caused by the flash is used to detect depth edge for non-photorealistic rendering [Raskar et al. 2004] or stereo vision [Feris et al. 2005]. In this paper, we use flash/no-flash images for a different application—matting.

3 Flash Matting

In this section, we describe the flash matting model and propose our joint Bayesian flash matting algorithm.

3.1 Flash matting model

For a static foreground and a fixed camera, we assume that the alpha channel of the foreground is unchanged in the no-flash image I (also called ambient image) and the flash image I^f. We have the matting equations for I and I^f:

$$I = \alpha F + (1-\alpha)B, \tag{2}$$

$$I^f = \alpha F^f + (1-\alpha)B^f, \tag{3}$$

where $\{F, B\}$ are the ambient foreground and background colors, and $\{F^f, B^f\}$ are the flash foreground and background colors, respectively. Treating the flash as a point light source with intensity L, the radiance E due to direct illumination from the flash on surface point P is

$$E = L \cdot \rho(\omega_i, \omega_o) \cdot r^{-2} \cdot \cos\theta, \tag{4}$$

where $\rho(\omega_i, \omega_o)$ is the surface BRDF (ω_i and ω_o are flash and view directions with respect to the local coordinate frame at P), r is the distance from the flash, and θ is the angle between the flash direction and the surface normal at P. This inverse square law explains why the flash intensity falls off quickly with distance r.

When the camera and flash unit are together and the background scene is distant from the camera, the intensity change of the background in flash and no-flash images will be small, $B^f \approx B$. Thus, we have the following *flash matting equation*:

$$I^f = \alpha F^f + (1-\alpha)B. \tag{5}$$

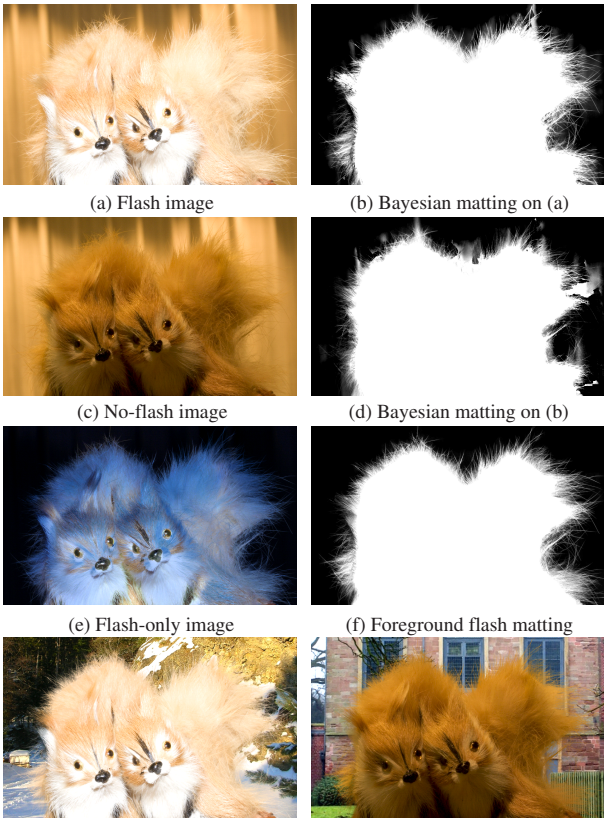
Figures 2(a) and 2(c) show a typical flash/no-flash image pair I^f and I. Notice that the background of the difference image $(I^f - I)$ (shown in Figure 2(e)) is dark, indicating that for this flash/no-flash pair, $B^f \approx B$.

Next we introduce our primary scenario for flash matting: *foreground flash matting*. Other extensions to flash matting will be described in Section 4.

3.2 Foreground flash matting

Subtracting (2) from (5), we have the *foreground flash matting equation*:

$$I' = I^f - I = \alpha(F^f - F) = \alpha F', \tag{6}$$



(a) Flash image (b) Bayesian matting on (a)

(c) No-flash image (d) Bayesian matting on (b)

(e) Flash-only image (f) Foreground flash matting

(g) Composition using flash image (h) Composition using no-flash image

Figure 2: Foreground flash matting. All results were generated using the same trimap.

where $F' = (F^f - F)$ is the additional flash foreground color. We refer to this difference image I' as the "flash-only" image in the paper. Figure 2(e) shows an example of a flash-only image.

Equation (6) is the foundation of our foreground flash matting because the flash-only image I' is independent of how similar the foreground and background are or the complexity of the background. However, the foreground flash matting problem is still under-constrained; to solve it, a trimap remains necessary. We will describe how to easily generate a good trimap from the flash-only image in Section 5. For now, we assume that the trimap is given.

Figures 2(a) and 2(c) show a flash/no-flash image pair I^f and I. We observe that the foreground and background colors are similar and the background is textured. It is very difficult to obtain high-quality matting result from either the flash image or the no-flash image, as shown in Figures 2(b) and 2(d). Figure 2(e) shows the difference image I' which clearly reveals the fine structures at the boundary of the foreground.

A straightforward approach to solving the foreground flash matting problem is to apply the Bayesian matting [Chuang et al. 2001] algorithm in two separate steps: 1) estimate α and F' from (6), and 2) use the recovered α, estimate F and B from (2) or F^f and B^f from (3).

In theory, this two-step approach yields a good alpha matte from (6) no matter how similar the foreground and background are and how complex the background. Figure 2(f) shows the computed alpha matte from image I' in the first step. The fine details of alpha matte were recovered well. Figures 2(g) and 2(h) show compositing results using estimated foreground colors from I^f and I in the second step.

In practice, however, the foreground flash matting equation (6) may be poorly conditioned when $\|F'\|$ is nearly zero. It is possible to get a dark flash-only foreground color F' if the foreground has low reflectivity ($\rho(\omega_i, \omega_o) \approx 0$), or if the surface normal is nearly perpendicular to the flash direction ($\theta \approx 90^\circ$). Another problem is pixel saturation due to highlights. Figure 3(e) shows the flash-only image of a toy lion. The fine white and red fur of the lion was not well lit by the flash. This resulted in an unsatisfactory foreground flash matting result (Figure 3(f)). While such ill-conditioned pixels may constitute a small fraction of all unknown pixels, humans are sensitive to incorrect local discontinuities.

Fortunately, these ill-conditioned pixels may be well-conditioned in the no-flash image I or flash image I^f. This observation motivates us to formulate joint Bayesian flash matting algorithm by fusing information from both images.

3.3 Joint Bayesian flash matting

For clarity, we present the basic idea of joint Bayesian flash matting using no-flash image I in (2) and flash-only image I' in (6). For each unknown pixel, we maximize a log likelihood function $L(\alpha, F, B, F'|I, I')$ of its unknown variables $\{\alpha, F, B, F'\}$, given the observation $\{I, I'\}$:

$$\begin{aligned} &\arg \max_{\alpha,F,B,F'} \quad L(\alpha, F, B, F'|I, I') \\ &= \arg \max_{\alpha,F,B,F'} \{L(I|\alpha, F, B) + L(I'|\alpha, F') + \\ &\qquad L(F) + L(B) + L(F') + L(\alpha)\}, \end{aligned} \tag{7}$$

where $L(\cdot)$ is the log of probability $P(\cdot)$. The term $L(I, I')$ is ignored because it is constant, and the log likelihood for alpha $L(\alpha)$ is assumed to be constant since we have no appropriate prior for a complicated alpha distribution.

The first two log likelihoods on the right hand side of (7) measure the fitness of solved variables $\{\alpha, F, B, F'\}$ with respect to matting Equations (2) and (6):

$$L(I|\alpha, F, B) = -||I - \alpha F - (1-\alpha)B||/\sigma_I^2, \tag{8}$$

$$L(I'|\alpha, F') = -||I' - \alpha F'||/\sigma_{I'}^2, \tag{9}$$

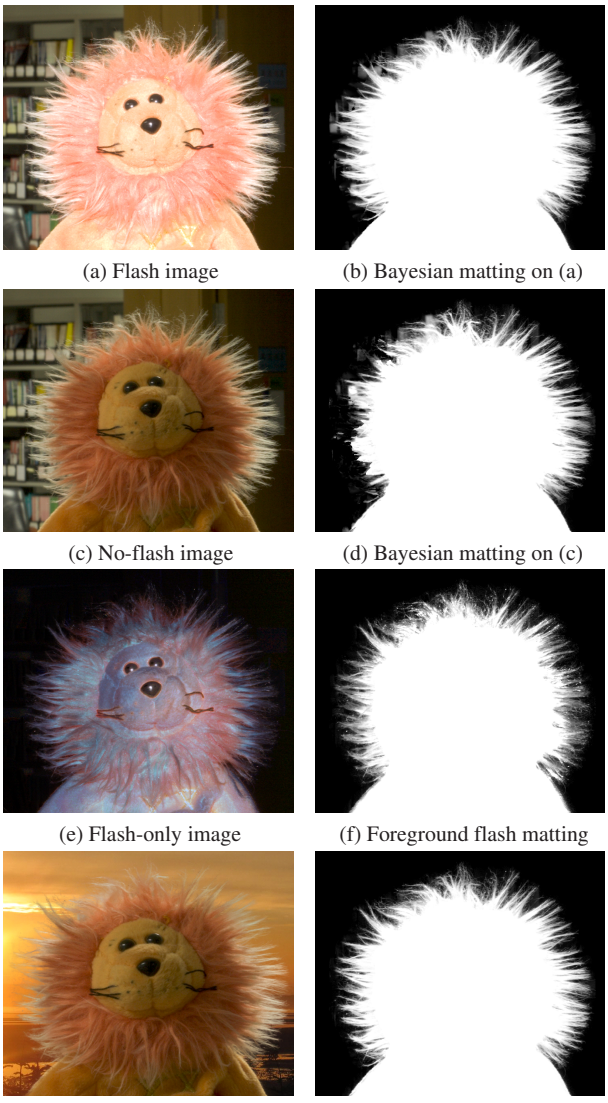
where σ_I^2 and $\sigma_{I'}^2$ are noise variances of images I and I' respectively. By default, we set these two variances to be the same in our experiments, with $\sigma_I^2 = \sigma_{I'}^2 = 32$.

We follow the color sampling technique proposed in Bayesian matting [Chuang et al. 2001]: a sliding window is used to sample neighborhood colors that contain previously computed values and unknown pixels are processed in a contour-by-contour scanning order. The statistics of foreground colors is represented as an oriented Gaussian distribution. The log likelihood $L(F)$ is modeled as

$$L(F) = -(F - \overline{F})^T \Sigma_F^{-1} (F - \overline{F}), \tag{10}$$

where $\{\overline{F}, \Sigma_F^{-1}\}$ are the mean and covariance matrix of the estimated Gaussian distribution, respectively. The background term $L(B)$ is defined in a similar way with $\{\overline{B}, \Sigma_B^{-1}\}$.

For the foreground color F', we also estimate an oriented Gaussian distribution $\{\overline{F'}, \Sigma_{F'}^{-1}\}$ in the flash-only image I'.



(a) Flash image (b) Bayesian matting on (a)
(c) No-flash image (d) Bayesian matting on (c)
(e) Flash-only image (f) Foreground flash matting
(g) New composition (h) Joint Bayesian flash matting

Figure 3: Joint Bayesian flash matting. Although foreground flash matting result is much better than Bayesian matting result on single image, there are still artifacts, especially at the top of the lion's head. Joint Bayesian flash matting produced the best alpha matte.

Thus,

$$L(F') = -(F' - \overline{F'})^T \Sigma_{F'}^{-1} (F' - \overline{F'}). \quad (11)$$

Taking the partial derivatives of (7) with respect to α and $\{F, B, F'\}$ and equating them to zero results in

$$\alpha = \frac{\sigma_{I'}^2 (F-B)^T (I-B) + \sigma_I^2 F'^T I'}{\sigma_{I'}^2 (F-B)^T (F-B) + \sigma_I^2 F'^T F'} \quad (12)$$

and

$$\begin{bmatrix} \Sigma_F^{-1} + \mathbf{I}\alpha^2/\sigma_I^2 & \mathbf{I}\alpha(1-\alpha)\sigma_I^2 & \mathbf{0} \\ \mathbf{I}\alpha(1-\alpha)\sigma_I^2 & \Sigma_B^{-1} + \mathbf{I}\alpha^2/\sigma_I^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{F'}^{-1} + \mathbf{I}\alpha^2/\sigma_{I'}^2 \end{bmatrix} \begin{bmatrix} F \\ B \\ F' \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_F^{-1}\overline{F} + I\alpha/\sigma_I^2 \\ \Sigma_B^{-1}\overline{B} + I(1-\alpha)/\sigma_I^2 \\ \Sigma_{F'}^{-1}\overline{F'} + I'\alpha/\sigma_{I'}^2 \end{bmatrix}, \quad (13)$$

where $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{0}$ the 3×3 zero matrix. To maximize (7), we iteratively estimate α and $\{F, B, F'\}$ using Equations (12) and (13) until changes between two successive iterations are negligible. At the beginning of optimization, $\{F, B, F'\}$ are set to $\{\overline{F}, \overline{B}, \overline{F'}\}$.

Note that (12) is not simple linear interpolation. It can adaptively select a well-conditioned matting equation from (2) or (6) to estimate the alpha matte α. If (2) is ill-conditioned (e.g., $F \approx B$), the alpha estimate will be dominated by (6), i.e., $\alpha \approx F'^T I' / F'^T F'$. Alternatively, if (6) is ill-conditioned (e.g., $F' \approx 0$), alpha will be automatically estimated by (2), i.e., $\alpha \approx (F-B)^T (I-B)/(F-B)^T (F-B)$. Thus, the underlying mechanism for our joint Bayesian flash matting selectively fuses information from two images, robustly producing high-quality matting results.

Figures 3(b) and 3(d) show results produced by Bayesian matting on the flash/no-flash image pair (a) and (c). In this example, Bayesian matting did not work as well on the left side of the image because of the complex background (bookshelf). In the foreground flash matting result (Figure 3(f)), significant details are lost due to underestimated alpha values. The highlight pixels at the top of the lion's head resulted in distinctly noticeable artifacts. Joint Bayesian flash matting produced the best result, as shown in Figure 3(h).

3.4 Practical issues

There are practical considerations associated with flash matting. Here we discuss the issues of ensuring the flash/no-flash images are exactly aligned and minimizing shadows cast by the flash.

Image acquisition. The basic capturing procedure is simple: focus on the foreground and lock the focal length, fix the shutter speed and aperture at appropriate values to take a no-flash image, and then turn on the flash to take a flash image. To obtain a high-quality matte, we have mounted a camera on a tripod to obtain pixel-aligned images. Handling dynamic scenes is beyond the scope of this paper. All images are acquired in raw format and then converted into an 8-bit linear TIFF format. We disable white balancing, gamma correction, and other non-linear operations in the raw conversion utility so that the two images are converted identically. Unless stated otherwise, we use a Canon EOS-20D digital camera and a Canon Speedlite 580EX flash.

To maximize change in the foreground between the flash and no-flash images, we recommend setting the camera exposure compensation to -1 stop so the no-flash image is slightly under-exposed. Before capturing the flash image, we set the flash exposure compensation to +1 stop and allow the camera's though-the-lens light metering to determine the best flash intensity.

Continuous shot. Current prosumer-level digital cameras are capable of continuous capture, typically between three and five frames per second. The capturing process can be adapted to take advantage of this feature. When capturing fine details of a live subject, such as human hair in an outdoor scene, we first activate the flash. Two images are then taken using the camera's "continuous mode." The first captured image is the flash image. The second one will be the no-flash image because the flash appears only instantaneously for the first image, and will be in the process of recharging. Figure 4 shows the results using this mode of capture. Our matting results are shown at the bottom row of

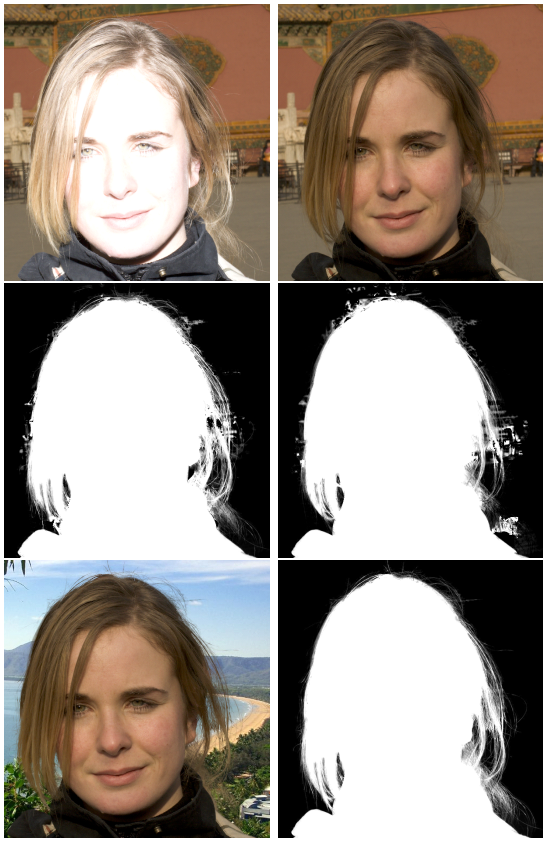


Figure 4: Continuous shot. Top: flash/no-flash pair. Middle: Bayesian matting results on the flash and no-flash images (separate estimation). Bottom: Joint Bayesian flash matting result.

Figure 4. Without flash matting, it is extremely difficult to extract a reasonably good matte for this example because of the complex background and similar foreground/background colors in certain areas.

Self-shadow. In flash photography, shadows are caused by two factors: depth discontinuities within the foreground object, and significant displacement between the flash unit and camera's optical center. Shadow pixels will erroneously result in $F' \approx 0$, thus degrading the flash matting results. If the foreground objects do not contain large internal depth discontinuities, the errors caused by shadow pixels are small and can be reduced by joint Bayesian flash matting. In fact, the flash/no-flash image pair shown in Figure 2 was captured using a consumer-level digital camera (Canon PowerShot G3) with its built-in flash. For a foreground object with large internal depth discontinuities, one solution is to use a ring-flash to produce a relatively shadow-free flash image. For example, the flash images in Figures 1 and 8 were taken using a Canon Macro Ring Lite MR-14EX ring-flash. The alternative is to use a normal external flash and capture multiple flash images by evenly varying the positions of the flash around the camera's center. The multi-flash option is detailed in Section 4.

4 Flash Matting Extensions

Foreground flash matting assumes a distant and static background, a relatively shadow-free flash image to achieve good results. Here we describe alternative flash/no-flash configurations that allow our joint Bayesian flash matting algorithm to work under general conditions.

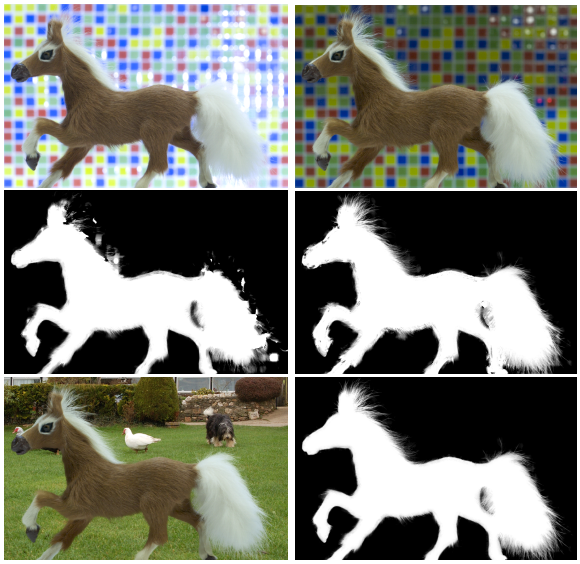


Figure 5: Background flash matting. Top: Flash/no-flash image pair. Middle: Bayesian matting results on the flash and no-flash images (separate estimation). Bottom: Recovered alpha matte and composition result. The flash was placed facing the background and away from the foreground object. It was triggered by a wireless flash transmitter.

Background flash matting. In scenes where the background and foreground are relatively close, the background change in the flash image may be too significant to be ignored. In such cases, we place the flash such that it illuminates the background and not the foreground. Assuming the foreground is virtually unaffected by the flash, as shown in Figure 5, we have

$$\begin{cases} I &= \alpha F + (1-\alpha)B \\ I^f &= \alpha F + (1-\alpha)B^f, \end{cases} \tag{14}$$

where we assume the foreground color is unaffected, yielding $F^f \approx F$. We now have the *background flash matting* equation:

$$I' = I^f - I = (1-\alpha)(B^f - B) = (1-\alpha)B', \tag{15}$$

where $B' = (B^f - B)$ is the additional flash background color. We can apply joint Bayesian flash matting again to maximize the likelihood $L(\alpha, F, B, B'|I, I')$. Figure 5 shows results using background flash matting.

Quick flash matting. In all previous flash/no-flash configurations, a fixed camera and a static background are assumed. However, a typical outdoor background scene may not be static. To handle such a condition, a more specialized flash unit is required. Current prosumer-level flash units, such as the Canon EX series, usually support a high-speed sync mode that allows the use of a higher shutter speed (e.g., 1/2000 sec) than camera's X-sync speed. (X-sync speed is the camera's top shutter speed that can be used with any flash. Typically, it varies from 1/60 sec to 1/250 sec).

The capture process is as follows. The no-flash image is captured under normal conditions to provide the correct colors. The flash image should be taken under *very short exposure*

Figure 6: Quick flash matting. Top: No-flash image and quick-flash image. The quick-flash image is taken with a shutter speed 1/2000 sec and an aperture F11. Middle: Bayesian matting results on the no-flash image and quick-flash image. Bottom: Joint Bayesian flash matting results (composition and extracted matte).

to ensure that the flash affects mostly the foreground object. We then get

$$\begin{cases} I &= \alpha F + (1-\alpha)B \\ I^f &= \alpha F^f \end{cases}, \tag{16}$$

where we assume that the background color $B^f \approx 0$ in the "quick-flash" image I^f. The likelihood $L(\alpha, F, B, F^f | I, I^f)$ is maximized by joint Bayesian flash matting; one such matting result is shown in Figure 6. In this example, it is very difficult to recover a high-quality matte from either no-flash image or quick-flash image. Notice that this process is not always applicable because it requires a relatively dark ambient illumination and a flash with high-speed sync mode.

Foreground multi-flash matting. To practically avoid shadows, we can use a ring-flash. The next best option is to capture a set of flash images $\{I_1^f, ..., I_N^f\}$ by evenly varying the position of the flash around the camera's center to approximate a ring-flash. This idea is very much in the spirit of prior multi-flash imaging work [Raskar et al. 2004; Feris et al. 2005]. From our experience, four images are sufficient (one each at the left, right, top, and bottom positions). A practically shadow-free flash image $\overline{I}^f$ is created using a pixel-wise maximum operation:

$$\overline{I}^f = \max\{I_1^f, ..., I_N^f\}. \tag{17}$$

Finally, the matte is computed by joint Bayesian flash matting. Figure 7 shows the shadow-free flash image created using four flash images, and the matting result.

5 Trimap generation

While matting techniques typically assume a user-supplied trimap, our method can produce a good initial trimap. We estimate the trimap from I' using a technique similar to Canny's two-pass method [1986] for edge detection. In the first pass, a global high threshold T is used to detect a foreground region Ω_F^1 with high confidence. We set T as the first local minimum of the histogram (128 bins) of the intensity of I'. The histogram is smoothed using a Gaussian kernel (with a variance of 7) to reduce noise. In the second pass, a lower threshold of $0.6T$ is used to detect foreground regions Ω_F^2 with lower confidence. Ω_F^2 is typically a set of connected components. We keep the components in Ω_F^2 that overlap with Ω_F^1 and remove all other isolated components in Ω_F^2. The second pass can effectively connect missed foreground regions from the first pass. Finally, the trimap is computed by dilating the boundary of $\Omega_F^1 \cup \Omega_F^2$. For objects with fur, the range of dilation is 5-20 pixels; for objects with solid boundaries, it is 2 pixels. Figure 8 shows the joint Bayesian flash matting result of a very complex plant using an automatically generated trimap.

Figure 7: Foreground multi-flash matting. Top: Flash image when flash is at left of the camera, and recovered alpha matte. Bottom: Shadow-free flash image generated from four multi-flash images, and recovered alpha matte.

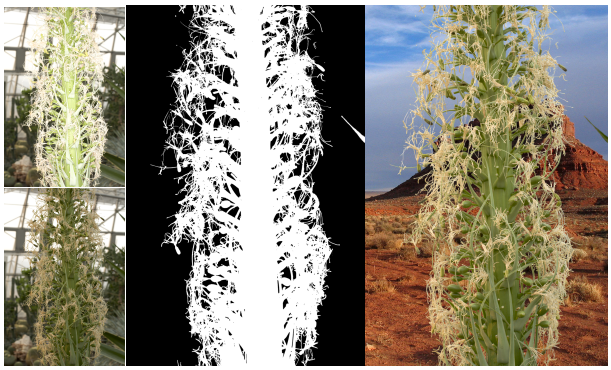


Figure 8: Results for a highly detailed plant. From left to right: Flash and no-flash image pairs, recovered alpha matte, and composition result.

In most of our experiments, the user needs to interactively adjust only the threshold T and dilation width to produce a reasonable-looking trimap. For very complex cases, we have developed a paint-style interface to touch up the trimap.

6 Limitations and Discussion

One limitation for foreground flash matting is the assumption that only the appearance of foreground is dramatically changed by the flash, due to the inverse square law. In practice, however, we have found that our joint Bayesian flash matting is robust enough to handle small appear-

Figure 9: A failure case. Top: flash/no-flash pair and initial trimap. Bottom: foreground flash matting result, joint Bayesian flash matting results, and a close-up view.

ance changes in the background caused by flash. To make our flash matting more robust to background change, an isotropic Gaussian distribution centered at zero can be used to model the ignored the flash-only background color B': $L(B') = -B'^T \Sigma_{B'}^{-1} B', \Sigma_{B'} = \sigma_{B'}^2 \mathbf{I}$, where $\sigma_{B'}^2$ is the variance to model the small change B'. The new log likelihood $L(\alpha, F, B, F', B'|I, I')$ is differentiable such that the iterative optimization still applicable.

As we discussed in Section 3.2, additional problems in flash matting include low reflectance foreground, surface near silhouettes, and pixel saturation. Figure 9 shows one such failure case. The black mug body results in errors in the trimap and under-estimated alpha values. As can be seen in the close-up view, the errors occurred near the silhouette and in saturated areas. The basic assumption of flash matting is violated because these areas were unaffected by the flash. However, even for this case, joint Bayesian flash matting still substantially outperformed foreground flash matting.

Our approach also assumes that the input image pair is pixel-aligned, which is common with many approaches that combine multiple images for enhancement (such as HDR and other flash photography techniques). Therefore, it is difficult to extract high-quality mattes in cases where fine structures move, as in the case of fur or hair photographed in a windy outdoor environment. One possible solution is to customize the capture system for (almost) simultaneous flash/no-flash capture, for example, using a two-camera system with a common virtual camera center and a beam-splitter [McGuire et al. 2005]. This system could also be electronically set up to capture flash/no-flash images in rapid succession. One camera is triggered slightly later than the other, with the delay being only the flash duration (typically 1 msec). The flash is activated so that only the first triggered camera records its effect. Another possibility would be to use a single programmable imaging camera [Nayar and Narasimhan 2002], with the exposure time of each pixel being independently controlled.

7 Conclusion

We have proposed a novel image matting approach using flash/no-flash pairs. Our flash matting approach capitalizes on the dominant change in the foreground appearance caused by flash to disambiguate the foreground and background. In particular, we have developed a joint Bayesian flash matting algorithm for high quality matte extraction even when the foreground and background are indistinguishable or the background has complex color distributions.

Flash matting is practical and easy to use. It does not require a special, controlled environment. It can even be used outdoors. Moreover, we have shown that the trimap can be quickly obtained even for highly complex objects. As a result, the average user of a camera equipped with a flash can easily use our flash matting technique.

Acknowledgements. We would like to thank the anonymous reviewers for their constructive critiques. Many thanks to Stephen Lin for his help in video production, and Dwight Daniels and Kurt Akeley for improving the manuscript.

References

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

Berman, A., Vlahos, P., and Dadourian, A. 2000. Comprehensive method for removing from an image the background surrounding a selected object. *U.S. Patent 6,134,345*.

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

Canny, J. 1986. A computational approach to edge detection. *IEEE Trans. on PAMI. 8*, 6, 679–698.

Chuang, Y.-Y., Curless, B., Salesin, D. H., and Szeliski, R. 2001. A bayesian approach to digital matting. In *Proceedings of CVPR 2001*, vol. II, 264–271.

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

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

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

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

Hillman, P., Hannah, J., and Renshaw, D. 2001. Alpha channel estimation in high resolution images and image sequences. In *Proceedings of CVPR 2001, Vol. I, 1063-1068.*

McGuire, M., Matusik, W., Pfister, H., Hughes, J. F., and Durand, F. 2005. Defocus video matting. In *Proceedings of ACM SIGGRAPH*, 567–576.

Nayar, S. K., and Narasimhan, S. G. 2002. Assorted pixels: Multi-sampled imaging with structural models. In *Proceedings of ECCV 2002*, vol. IV, 636–652.

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

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

Ruzon, M. A., and Tomasi, C. 2000. Alpha estimation in natural images. In *Proceedings of CVPR 2000*, 18–25.

Smith, A. R., and Blinn, J. F. 1996. Blue screen matting. In *Proceedings of ACM SIGGRAPH 1996*, 259–268.

Sun, J., Jia, J., Tang, C.-K., and Shum, H.-Y. 2004. Poisson matting. In *Proceedings of ACM SIGGRAPH*, 315–321.

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

Wang, J., and Cohen, M. 2005. An iterative optimization approach for unified image segmentation and matting. In *Proceedings of ICCV 2005*, vol. II, 936–943.