Eurographics Symposium on Rendering (2006)
Tomas Akenine-Möller and Wolfgang Heidrich (Editors)

Practical, Real-time Studio Matting using Dual Imagers

M. McGuire,[1] W. Matusik,[2] and W. Yerazunis[2]

[1]Williams College, Williamstown, MA
[2]Mitsubishi Electric Research Laboratory, Cambridge, MA

Abstract

This paper presents a practical system for capturing high-resolution video mattes using cameras that contain two imagers on one optical axis. The dual imagers capture registered frames that differ only by defocus or polarization at pixels corresponding to special background 'gray-screens.' This system eliminates color spill and other drawbacks of blue-screen matting while preserving many of its desirable properties (e.g., unassisted, real-time, natural illumination) over more recent methods, and achieving higher precision output for Bayer-filter digital cameras. Because two imagers capture more information than one, we are able to automatically process scenes that would require manual retouching with blue-screen matting.

The dual-imager system successfully pulls mattes for scenes containing thin hair, liquids, glass, and reflective objects; mirror reflections produce incorrect results. We show result comparisons for these scenes against blue-screen matting and describe materials and patterns for building a capture system.

Categories and Subject Descriptors (according to ACM CCS): I.4.6 [Segmentation]: Pixel classification

1. Introduction

Matting is the problem of separating an input image I into three output images: a background B, foreground F, and *matte* α that represents the sub-pixel coverage of the background by the foreground at each pixel. The background is typically discarded and the matte used to composite the foreground into a new scene. This process is used for special effects shots and is seen nightly in news programs that superimpose a weather reporter over a satellite map.

We describe a new, practical method for everyday matting in film and television studios that has the following desirable properties:

- Extremely precise sub-pixel results
- Natural illumination
- Regular camera form-factor
- Works with many synthetic and organic materials
- Robust to illumination changes and shadows
- No color reflected onto foreground
- No tinting of thin features
- Real-time matte extraction (1280 × 960 at 30fps)

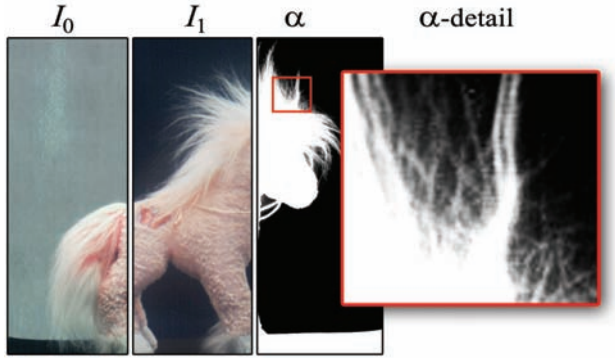

Figure 1: *Input images $I_{0,1}$ that differ in polarization; a high-resolution output matte α pulled by our system.*

The precision comes from using all color channels to pull the matte (instead of just blue or green), from achieving near-perfect registration of the sensors (compared to previous multi-sensor methods [ZKU*04,MMP*05]), and from the high contrast avail-

able on our prepared background screens. We are unaware of another system capable of autonomously pulling mattes at such high resolution; at the end of this paper in Figure 9 we show individual sub-pixel hairs accurately matted at a distance of several meters from the camera.

Using one camera with dual imagers, our method captures simultaneous frames I_0 and I_1 that differ by about half the dynamic range at background pixels and are identical at foreground pixels, so that $\alpha \approx 1 - 2(|I_0| - |I_1|)$. We built several cameras that each capture video with this property. Each design uses one of two properties–defocus or polarization–and works with a specifically chosen neutral-color background screen. Like blue-screen matting, the dual-imager is limited to scenes that do not contain mirror reflections (although most metals are fine) and requires the use of a background screen. One advantage of blue-screen matting that we do not support is that blue-screen can mask flexible foreground objects, e.g., a blue glove causes an actor's hand to intentionally disappear. Our polarization screen is too inflexible to use as cloth for this purpose and defocus depends on a depth discrepancy, which makes matting out foreground objects impossible.

Shiny objects tend to produce polarized specular highlights at glancing angles, which theoretically limits our polarizer system when working with those. However, the fact that those highlights saturate corresponding pixels turns out to be more of a problem than polarization. In practice we observe that even a polarized reflection has a smaller contrast ratio than the black-to-gray transition of our background screen, so polarized reflections do not appear to be a problem. To demonstrate this we show results on metals, glass, and other specular objects.

Our contribution is extending and improving previous methods for studio matting. We bring them to maturity through new research and systems work so that they can be practically used in real applications. These include film production, video conferencing, and augmented reality. Specifically, we describe a theoretically-justified algorithm for sub-pixel coverage that accounts for noise and other imperfections present in real imaging systems; report on a new fast time-multiplexing camera for defocus and polarization and a new behind-the-lens polarization camera; and have created a small database of matting test sequences. The previous lack of standard test sequences has made result comparison challenging in the literature and impeded research and educational exploration of matting.

2. Related Work

Matting from a single image is underconstrained [SB96], so it is necessary to obtain more information by various methods. *Assisted* methods [CCSS01, HHR01, RKB04, RT00, WBC*05, LSS05, WC05] refine coarse user-painted mattes. These are intended as editing tools, not for real-time or long videos, although they have been extended by stereo disparity [ZKU*04] and keyframe interpolation [CAC*02].

Active methods shine imperceptible (IR [DWT*02, YI02, Vid60, YNH04], UV, polarized [BE00], sodium [Vla58], microstrobes [WGT*05]) light to capture two images that differ at the background. Although these produce high-quality mattes, they are only used occasionally in production because regular studio lighting interferes with active illumination and common materials have varying responses to imperceptible light (e.g., as shown for different kinds of black cloth by Debevec et al. [DWT*02]). One active system of particular note is Reflecmedia's (http://reflecmedia.com) active blue-screen. It pairs a ring of low-intensity colored LEDs around the camera lens with a retro-reflective, gray screen. Light from the LEDs reflects directly back to the camera, creating the appearance of a blue background for traditional blue-screen matting, but is of sufficiently low intensity that minimal blue-spill occurs. The company's website reports an "immediate key that is 90% towards the finished result," typically followed by a manual touch up, as well as a separate real-time preview product for low-quality results. Unfortunately, no results suitable for comparison are publically available. We speculate that for scenes on which blue-screen matting works, their high-quality, manual results are better than our fully automatic ones but have not had an opportunity to experiment with their system.

Passive methods work with natural, visible (as opposed to coherent, structured, or computer-controlled) illumination. They compare a primary video stream to auxilliary ones that are previously known (background subtraction [QS99], environment matting [ZWCS99, CZH*00], triangulation [SB96]), constant color (chroma-key [Mis92], luminance matting, blue-screen [SB96, Vla71]), or defocussed [MMP*05]. Passive methods are easier to integrate into the film production pipeline because they work with existing infrastructure and are more robust to the reflectivity of scene materials. Another benefit of passive methods is that they can be used for video-conferencing because they are not distracting to the actor in the way that, for example, time-multiplexed backlighting can be.

Our system is passive. The closest active method to ours is Ben-Ezra's invisible key segmentation [BE00], which illuminates the scene with polarized light, or alternatively employs a polarized back-light, and segments the image based on polarization with a chroma-key-like algorithm. We extend their beam-splitter camera design with new alternatives and produce high-resolution video results. We believe that our extension is more practical for studio use because it works with natural illumination and produces accurate fractional α values; the latter of which we attribute to a more general algorithm based on partial-coverage composition [PD84] instead of segmentation.

The dual-imager approach is similar to two passive methods. Our core algorithm is mathematically identical to triangulation matting [Wal82, SB96], which was restricted to images of static scenes. The most significant difference is that we show how to capture the two images simultaneously and how to remove the background terms, which allows us to operate on video. For video, we need a more robust algorithm than triangulation matting because we cannot afford to micro-adjust constants every frame, so we extend the core with new terms. Defocus matting [MMP*05] captures three differently-focussed video streams of scenes with arbitrary backgrounds and pulls the matte via optimization in minutes per frame at 320×240. We require a special background, but produce much higher-resolution mattes from a normal form-factor camera, and do so in real-time.

3. Algorithm

Let I_0 and I_1 be images of the same pre-multiplied foreground αF against backgrounds B_0 and B_1, given by [PD84]

$$I_i = \alpha F + (1 - \alpha)B_i. \tag{1}$$

Smith and Blinn [SB96] solve for the matte at each pixel as

$$\alpha_T = 1 - [I_0 - I_1]/[B_0 - B_1], \tag{2}$$

where $[\cdot]$ denotes luminance (mean of RGB channels.) They assume that B_0 and B_1 are known and different at all pixels. For video, B_0 and B_1 are unknown because they change as the camera moves and the actor casts shadows.

We introduce user-controlled parameters $\delta \approx [|B_0 - B_1|]$ and $b_0 \approx [B_0]$ that approximately describe the unknown background images. Assume that the scene satisfies the *uniformity property*: **δ and b_0 are constant over the image even if B_0 and B_1 vary** (we show how capture images with this property in a moment).

Given δ and b_0, we no longer need explicit B_i images and can pull a matte from the I_i alone. For robustness, we combine triangulation (α_T) with conservative luma (α_L) and saturation (α_S) mattes,

$$\alpha'_T = 1 - |[I_0 - I_1]|/\delta \tag{3}$$

$$\alpha_L = |[I_0] - b_0| * k_1 \tag{4}$$

$$\alpha_S = [|I_0 - b_0|] * k_4 \tag{5}$$

$$\alpha = \max(\alpha_T, \alpha_L, \alpha_S) * k_2 + k_3. \tag{6}$$

Note that b_0 is scalar; when we mix colors and scalars in an equation the scalar spreads to an RGB triplet, e.g., (b_0, b_0, b_0).

In eq. 3 we are able to use the absolute value because we know that $\alpha \leq 1$. Several constants are available for tuning: δ and b_0 describe the exposure and contrast of the image, k_1 is the standard luminance matte control and k_4 its saturation equivalent, and k_2 and k_3 are contrast and bias enhancement for the matte. These operator-specified constants correspond to equivalent constants used in blue-screen matting and postproduction cleanup of mattes (see [SB96]; for convenience we follow their k-notation). We give useful values for these constants in Section 5. The most sensitive is δ. Figure 2 shows that improper settings lead to loss of small features (when δ is too low) and noise (when δ is too high).

Given α, we solve eq. 1 for the pre-multiplied foreground,

$$\alpha F = I_0 + b_0(\alpha - 1). \tag{7}$$

Smith and Blinn photographed I_0 and I_1 against two background screens at different times. By varying aperture or filters between two imagers, we can simultaneously obtain different images from a single background screen. We now describe two such scenarios that satisfy the uniformity property, and for each extend the dual-imager solution to take further advantage of those scenarios.

3.1. Defocus

Let the true (unknown) background B be a high-frequency black and white noise pattern with mean $\frac{1}{2}$, I_0 be the image formed by a wide aperture focussed away from the background, and I_1 be the image

Figure 2: *Effect of the δ parameter. Mattes at $\delta = 0.18, 0.34, 0.46$.*

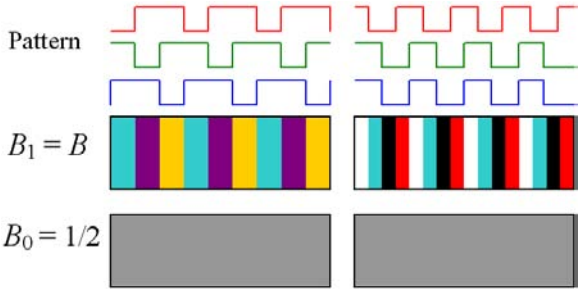

Figure 3: Top: *Two defocus patterns.* Bottom: *Their pinhole and defocussed images. The pattern on the left is theoretically optimal, but in practice blue and green are often poorly distinguished by digital cameras so we prefer the pattern on the right.*

formed by pinhole. Defocus blurs B on the first imager, so b_0 is uniform gray. Because B contains only black and white, the absolute difference between B_0 and B_1 must be uniform $\delta = \frac{1}{2}$ if the background pattern is aligned with the imager pixels. Note that the process just described is not defocussing by changing the distance between the lens and the imager, but by holding that distance fixed and changing the aperture (and therefore, the point spread function) radius.

We need not use noise for the background screen; vertical bars are sufficient as long as the period is at most half the point-spread radius of the lens and each bar is at least one pixel wide. Black-and-white bars provide the necessary high frequencies. However, since the background pattern will typically not be perfectly aligned with the sensor pixels, a monochrome pattern is insufficient. Pixels that sample the edge between black and white will appear gray and decrease the effective contrast ratio. To address this, Figure 3 (left) shows a theoretically optimal pattern, where strong edges are present in every color channel and no pinhole sensor pixel will measure gray.

In practice, it is difficult to exactly match illumination spectra and the reflectivity of dyes in the background screen to the spectral response of the camera. This problem is compounded by the mismatch between the gamut of the printing process used to create the screen and the camera's sensors. We found that in practice printed green and blue are particularly hard to distinguish on a camera, especially under fluorescent lights (the typical case for video conferencing). Figure 6 shows that the sensor's blue and green responses share significant overlap, even under ideal full-spectrum illumination. Therefore we created and actually use the pattern shown in Figure 3 (right) that has strong luminance edges on all channels and no edges that appear gray, but anticipates correlation of blue and green and always varies them synchronously.

We can exploit the regularity of these patterns. When the camera moves it is straightforward to register the background orientation and determine the in-focus B_0, so the background is always known. This allows two enhancements. First, we can extend eq. 7 to blend images from both cameras:

$$\alpha F \;=\; \big((I_0 + I_1) + (\alpha - 1)(b_0 + B_1)\big)/2. \qquad (8)$$

Second, we can pull α_T using eq. 2, although instead of luminance, we compute separate mattes for each color channel and choose the one with the best conditioned denominator at each pixel. We interpolate pixels where no channel is well-conditioned from neighbors using the push-pull algorithm [GGSC96].

3.2. Polarization

Let the background be a white screen laminated with a horizontal polarizing filter, I_0 be horizontally polarized, and I_1 be vertically polarized. Under natural illumination, half the light reflected by the screen reaches imager zero and none reaches imager one. Therefore B_1 is mostly gray, B_0 is mostly black ($b_0 = 0$), and everywhere their difference is $\delta \approx \frac{1}{2}$ so uniformity is satisfied and eq. 7 is applicable.

We can do even better: the background appears black in I_1, so the additional conservative luminance matte estimate $\alpha_L = I_1 k_1$ is available for the max operation. We can also extend eq. 7 by

$$\alpha F \;=\; \big((I_0 + I_1) + (\alpha - 1)(b_0 + b_1)\big)/2, \qquad (9)$$

where b_i is user-controlled intensity estimate of B_i. Although from our derivations, $\delta = |b_0 - b_1|$, in practice it is useful to set these independently; b_i limits darkening (i.e., color bleeding of the background) at edges and δ the overall discrimination (see Figure 2).

Noise may appear in the background areas of the matte when the polarization contrast ratio is poor[†]. To suppress this, we employ a form of background subtraction. Identify dark ($|I_1| < 0.2$), desaturated ($\big[I_0 - |I_0|\big] < 0.2$) areas that retain *some* contrast ($[I_1]{*}1.5 < [I_0]$) and diminish α there by $([I_0]-[I_1])/2$.

4. Capture Systems

There are many ways of building a camera with dual-imagers that share an optical axis: a beam-splitter in front of the lens [Wol94, DWT*02, BE00,

[†] Rotating the imager any angle θ about the optical axis gives $\delta \approx |\cos(2\theta)|/2$, so 90° produces the optimal contrast ratio.

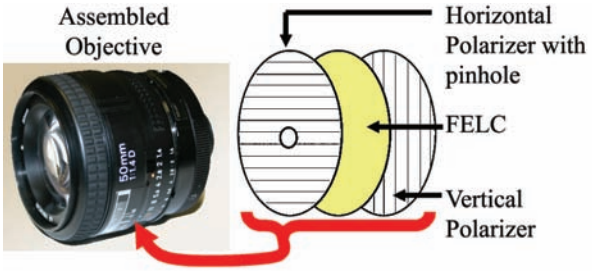

Figure 4: *Photograph of an objective containing the high-speed iris FELC diagrammed on the right.*

MMP*05], a beam-splitter behind the lens, time-multiplexing [WGT*05], per-pixel polarization filters [BE00, WMPA97], and a refocussing light field camera [NLB*05]. We experimented with the first three.

4.1. Time Multiplexing

Time multiplexing involves changing the polarization filter or lens aperture at 60 Hz to capture two time-interlaced 30 fps video streams. Because only one physical imager is used, the images are perfectly aligned in color and space. Of course, they are not aligned in time, so it is desirable to both capture at high frame rate, say 1 kHz, and correct remaining displacement errors with optical flow (see [WGT*05]). Changing lens parameters at 1 kHz is a challenge. No physical iris or filter can move at that speed without its momentum vibrating (or destroying!) the camera, and LCD apertures have approximately 1/60s response times.

For fast focussing, we created the new electronic iris shown in Figure 4, which fits into a normal camera objective ('lens'). It contains perpendicular polarizers that together block all incident light, except at a pinhole in the first polarizer. Between these we position a plate of ferroelectric liquid crystal (FELC). This is an active optical component that rotates the polarization of incident light 90° when voltage is applied across it; doing so expands our pinhole to a wide aperture. We drive our iris from the camera's hardware trigger line. FELC is much faster than LCD and can switch at up to 10 kHz; faster than our high-speed QImaging Retiga 1300 camera.

The same technology can be applied to rapidly switching polarization. For that application we place a single polarizer between the FELC and the imager. Toggling the FELC thus alternately allows horizontally- and vertically-polarized light from the scene to reach the imager.

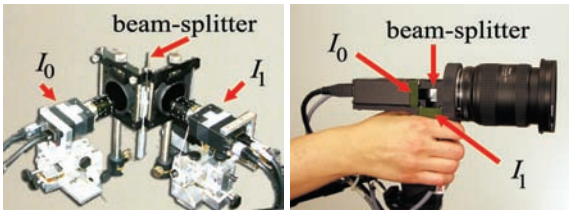

Figure 5: *Our cameras that split in front of (left) and behind (right) the lens; the latter allows a smaller, hand-held form factor.*

4.2. Beam-Splitters

Beam-splitters are glass plates or cubes with an internal surface that partition an incident light field between two perpendicular output paths. This produces two images that are perfect synchronized in time; the challenge is aligning them in color and space.

For defocused imagers, we use a plate splitter and two lenses, as shown in Figure 5 (left). This allows different apertures on each camera. For polarizing imagers, we can either use the same design or move the beam splitter behind the lens, as shown in Figure 5 (right). One can achieve different polarizations using filters, however a broadband polarizing cube beamsplitter (like the Melles Griot 03PBB002) maintains near-ideal efficiency by selectively reflecting light based on polarization.

After several unsuccessful attempts to manufacture aluminum parts to precisely position the imagers, we discovered a neat trick: creating a plastic camera housing directly from 3D geometry using a Dimension SST 3D printer.

4.3. Calibration

Spatial calibration is a one-time part of the camera construction process. The imagers need only be perfectly aligned at the plane of focus where the actor will stand because B_0 is always uniformly gray and never appears misaligned compared to B_1. This means that we only need to solve a 2D calibration problem. Assuming negligible lens distortion, a 2D affine mapping is sufficient to register I_1 to I_0. We solve for this as a least squares problem on corresponding feature points sampled from a checkerboard. Given 2D point sets v_0 and v_1 expressed as matrices whose columns have the form $[x\ y\ 1]^T$, we seek the 3×3 matrix H that minimizes $|Hv_0 - v_1|^2$. The solution requires a pseudo-inverse that tends to be ill-conditioned; as is common practice, we solve for it using singular value decomposition with small values ignored along the diagonal.

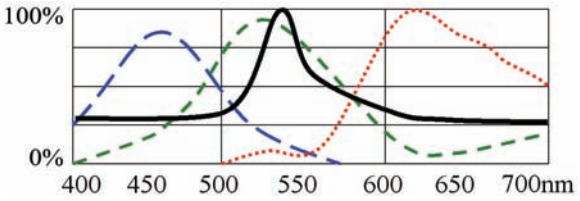

Figure 6: *Normalized RGB response of our cameras (dashed) and the reflectivity of green-screen fabric (solid). Note that the green-screen correctly peaks near our sensor's green response, and that the sensor's blue and green responses share substantial overlap.*

We use Basler A631fc cameras with an automatic white-balance control to correct for color discrepancies between the cameras. We block IR and UV wavelengths with Tiffen 77SHM and 77UVP filters; without these there is a brightness offset.

4.4. Screen Construction

We printed the defocus screens on a large-format printer with diffuse paper. Beware that the printer and camera gamut will differ, so it is necessary to screen-test for the particular combination.

We constructed polarizing screens by laminating museum-quality 'art rag,' an extremely white and diffuse paper made from cotton fiber, with inexpensive self-adhesive sheet polarizers (e.g., #POAT from 3Dlens.com), which are produced for industrial applications. The tendency of the sheets to produce specular reflections can be mitigated by sanding them uniformly, which does not significantly reduce their diffuse reflectivity.

5. Results

We pulled mattes for hard test scenes using several systems, and here report typical positive and negative results for the best systems. We adjusted parameters to give best results for both algorithms. The dual-imager constants were typically around $b_0 = 0.6$, $k_1 = k_4 = 2.0$, $k_2 = 2.0$, $k_3 = -0.4$, and $\delta = 0.2$. For defocussed imagers we assumed perfect background registration and used the B_i images on the right in Figure 9.

Figure 8 compares polarized, split-behind-the-lens dual-imagers to blue-(actually, green-)screen matting, which is the best prior solution for high-resolution mattes in real-time from natural illumination. The green screen was Rosco DigiComp matting fabric. This is the same professional-quality material that is used every day in film studios for matting. Figure 6 demonstrates that the fabric's spectral reflectivity peaks near the green sensor response peak, which confirms that we built a fair test case for green-screen matting.

In all tests we used full-spectrum illumination for maximum color discrimination and attempted to light the background as uniformly as possible. In the course of film postproduction it is common practice to manually retouch and filter mattes to remove blue-screen artifacts. We did not do so in these tests for two reasons. First, an objective comparison of matting methods is impossible with human retouching. Second, our goal is robust, fully-automated matting and postproduction is user assistance – assistance that would not be available for other applications like video conferencing, augmented reality, and machine vision. We acknowledge that films will always use postproduction to achieve the highest quality result and endeavor to minimize the artifacts that studios need to correct.

Each row of the Figure 8 shows the input frames I_0 and I_1, the recovered matte α, a novel composite using that matte, and details of particularly good or bad results .

Rows 1 and 2 of Figure 8 are a hard test case, with thin fur, reflective metal objects, many colors, a mirror, and a translucent glass. Figure 7 shows enlarged images of the output mattes. The dual-imager fails on the mirror, although not as badly as blue-screen does, but correctly pulls a matte for all colors and even for translucent and reflective objects. The detail images show that our mattes are higher resolution than single-imager blue-screen matting, capturing fine details like the fur. We attribute this increased resolution to the fact that our matte is based on saturation and luminance, which are sampled at each pixel. For a Bayer-patterned imager, hue is sampled at 1/4 (for blue) or 1/2 (green) that resolution, which unsurprisingly leads to proportionally more coarse results for the blue-screen algorithm.

Green-screen of course fails for green objects. This prevents matting of vegetation, as shown in row 3. The "white" table has a green hue from reflected light off the background; we have set the green-screen constants artificially loose to demonstrate that this problem really can occur (the vegetation becomes translucent no matter how tight we set the control constants.) We could ask the weather forecaster in row 5 to change her green shirt, but she is unlikely to change her green eyes. Note the green spill on the hand in the detail image for row 5, a classic artifact from that algorithm. Dual-imagers perform well in these cases.

The detail in row 6 shows a strong result: individual hairs matted by dual imagers; and an artifact: the white backing visible at the seam between two polarization screens appears as a thin horizontal line in the

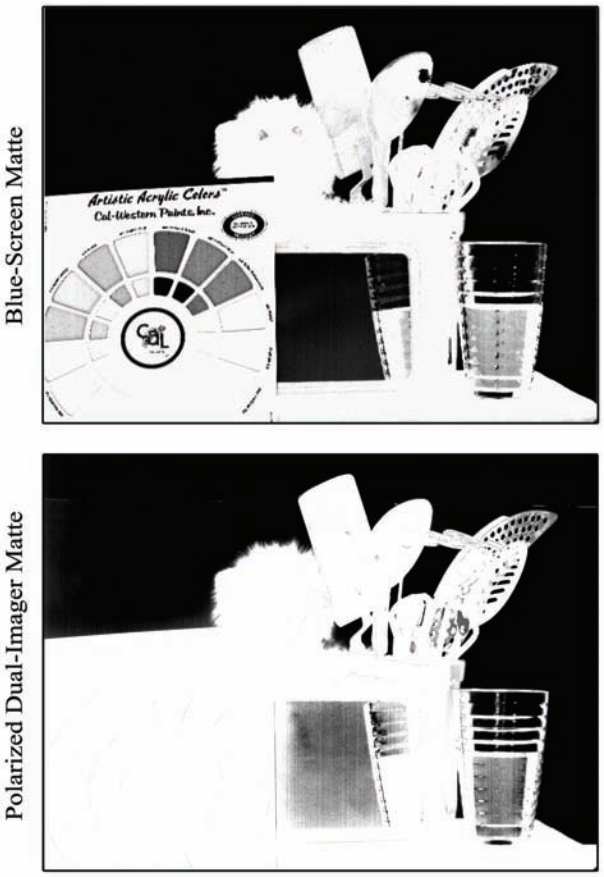

Figure 7: *Results on a hard test case with fur, metals, glass, mirrors, and colors.*

matte. We've also observed artifacts at specular highlights, which can appear much brighter in one imager than another if they are polarized. However, not all specularities are a problem, e.g., rows 2 and 4 show correct highlights on the spoons and bowl. One solution that we employ is to specifically set $\alpha = 1$ at saturated pixels; this is successful as long as specularities are not so bright that the darker background is also over-exposed.

Figure 9 shows results obtained with time-multiplexed, defocussed imagers that demonstrate the high precision of our results. These are 1 megapixel images; zooming in reveals hairs correctly matted with fractional α values. This is an extremely challenging case for matting; row 2 even shows a single sub-pixel hair detected from the angel, which is barely visible in the original photographs because it matches the background color.

Figure 10 compares blue-screen and dual-imager polarized matting for a moving liquid, in this case water with red food coloring poured from a measuring cup marked with gradations. On the left, blue-screen is unable to capture colored translucency well and produces a tinted foreground. Pixels of the "red" liquid near the arrow exhibit a greenish hue, with only a 8 7 red to green ratio. On the right, polarized dual-imagers capture smooth alpha transitions to produce good specular highlights and correct color, even though the Fresnel reflections off water are known to be polarized and could present a challenge. The dual-imager system produces a 2 1 red to green ratio for the red liquid.

6. Discussion

From working with several prototypes, we conclude that time multiplexing is well-suited to defocus because it gives a normal camera form-factor and requires no calibration. For polarization we prefer a beam-splitter behind the lens, to avoid optical flow and provide easier calibration than splitting in front of the lens.

Our approach is practical for studio use. The dual-imager algorithm is fast, our preferred cameras have regular form-factors, we produce sharper mattes than previous methods and our new screens are a straightforward replacement for blue and green ones.

What about outside the studio? Efficiently pulling unassisted mattes there is an open problem. Preliminary experiments indicate that some high-frequency but non-patterned backgrounds like trees and stucco walls are acceptable substitutes for our defocus screen. We plan to investigate background registration and inpainting techniques that can extend dual-imager matting for natural scenes.

Acknowledgement

Special thanks to Cliff Forlines for editing our video figures.

References

[BE00] Ben-Ezra M.: Segmentation with invisible keying signal. In *IEEE CVPR* (2000), pp. 32–37.

[CAC*02] Chuang Y.-Y., Agarwala A., Curless B., Salesin D. H., Szeliski R.: Video matting of complex scenes. *ACM Trans. on Graphics 21*, 3 (July 2002), 243–248.

[CCSS01] Chuang Y.-Y., Curless B., Salesin D. H., Szeliski R.: A bayesian approach to digital matting. In *IEEE CVPR 2001* (December 2001), vol. 2, IEEE Computer Society, pp. 264–271.

[CZH*00] Chuang Y.-Y., Zongker D. E., Hindorff J., Curless B., Salesin D. H., Szeliski R.: Environment matting extensions: towards

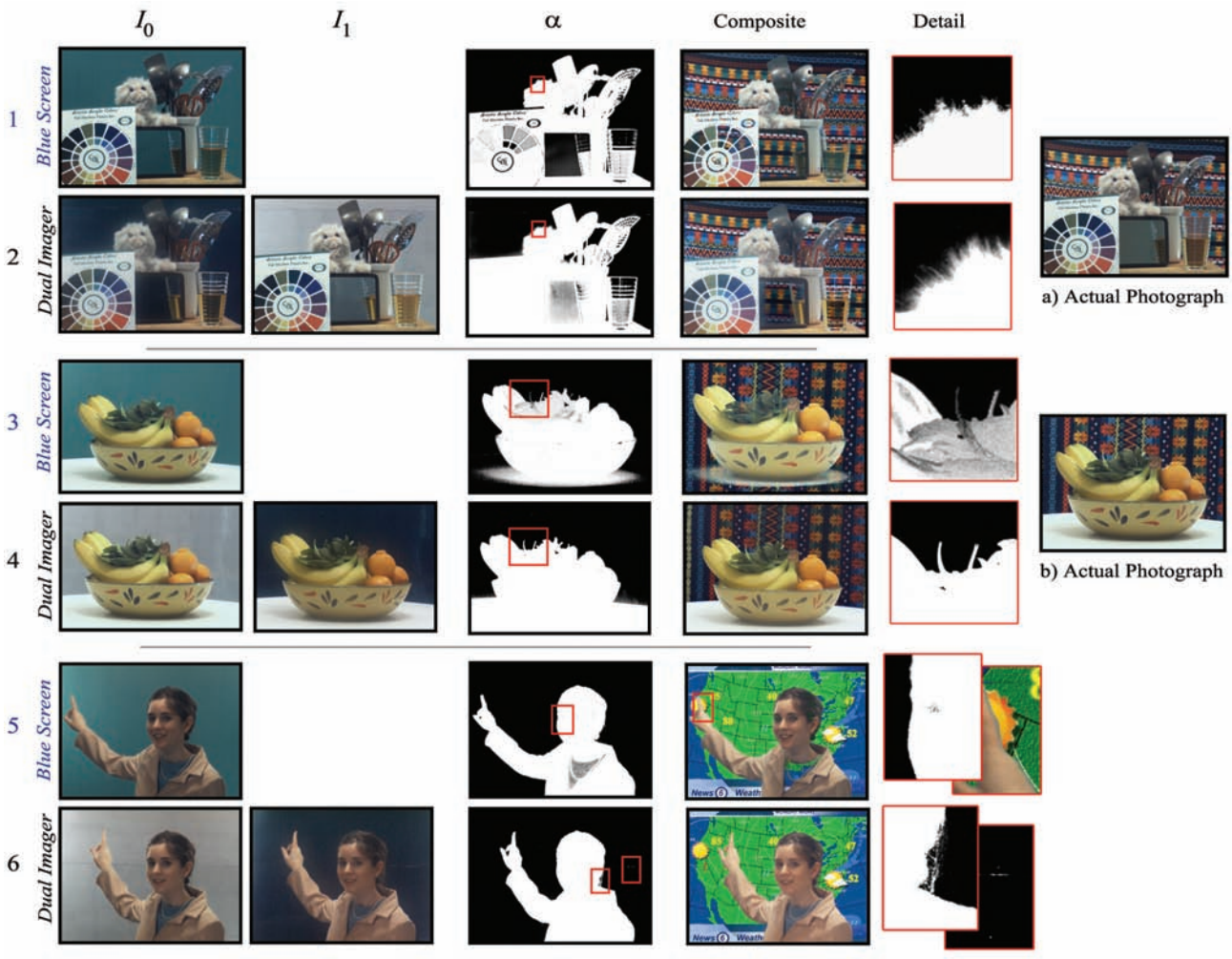

Figure 8: *Comparision of mattes pulled with traditional blue-screen versus polarized dual imagers. The rightmost column shows actual photographs as ground truth 'composites' (not applicable for the live weather forecast).*

higher accuracy and real-time capture. In *SIGGRAPH '00* (2000), ACM Press/Addison-Wesley Publishing Co., pp. 121–130.

[DWT*02] Debevec P., Wenger A., Tchou C., Gardner A., Waese J., Hawkins T.: A lighting reproduction approach to live-action compositing. *ACM Trans. on Graphics 21*, 3 (July 2002), 547–556.

[GGSC96] Gortler S. J., Grzeszczuk R., Szeliski R., Cohen M. F.: The lumigraph. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM Press, pp. 43–54.

[HHR01] Hillman P., Hannah J., Renshaw D.: Alpha channel estimation in high resolution images and image sequences. In *Proceedings of IEEE CVPR 2001* (December 2001), vol. 1, IEEE Computer Society, pp. 1063–1068.

[LSS05] Li Y., Sun J., Shum H.-Y.: Video object cut and paste. *ACM Trans. Graph. 24*, 3 (2005), 595–600.

[Mis92] Mishima Y.: A software chromakeyer using polyhedric slice. In *Proceedings of NICOGRAPH 92 (Japanese)* (1992), pp. 44–52.

[MMP*05] McGuire M., Matusik W., Pfister H., Hughes J. F., Durand F.: Defocus video matting. *ACM Trans. Graph. 24*, 3 (2005), 567–576.

[NLB*05] Ng R., Levoy M., Brédif M., Duval G., Horowitz M., Hanrahan P.: *Light Field Pho-*

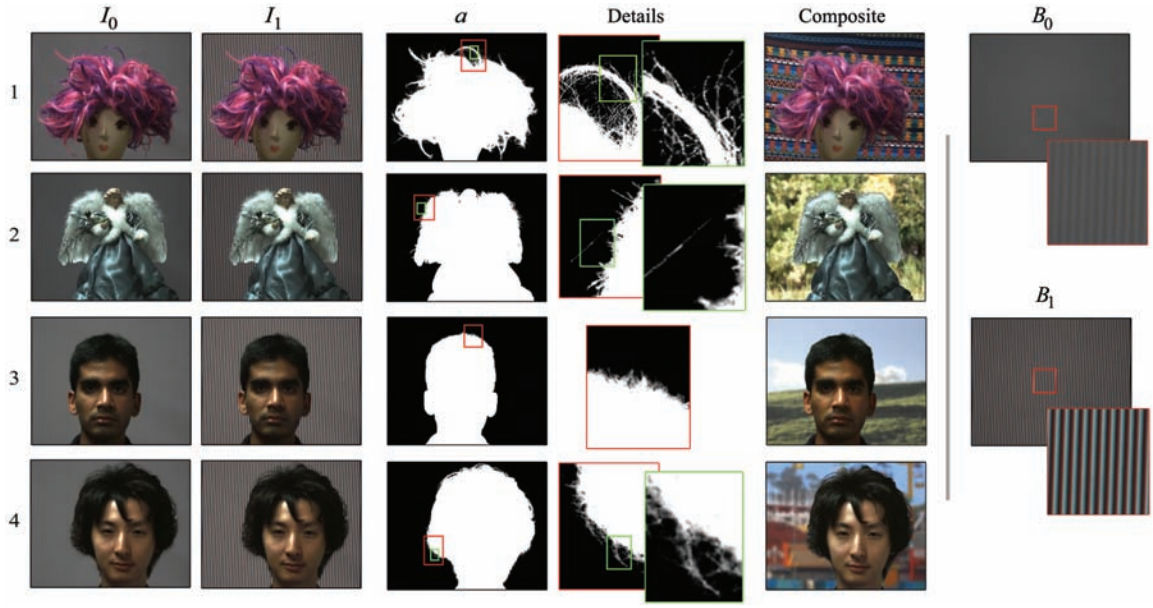

Figure 9: *Demonstration of extremely high-resolution mattes pulled from defocussed dual imagers.*

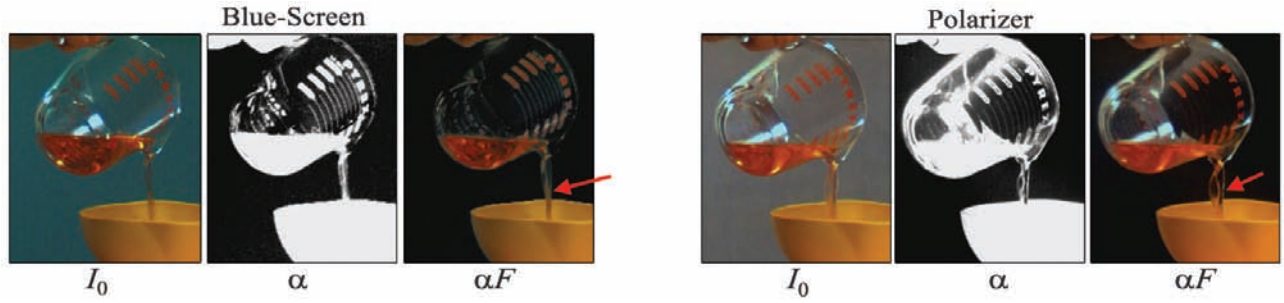

Figure 10: *Moving, translucent liquid. Polarized dual-imagers, as shown on the right, are better at extracting the color and matte.*

tography with a Hand-Held Plenoptic Camera. Tech. Rep. Tech Report CSTR 2005-02, April 2005.

[PD84] Porter T., Duff T.: Compositing digital images. In *SIGGRAPH '84* (1984), ACM Press, pp. 253–259.

[QS99] Qian R. J., Sezan M. I.: Video background replacement without a blue screen. In *Proceedings of ICIP* (1999), vol. 4, pp. 143–146.

[RKB04] Rother C., Kolmogorov V., Blake A.: "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics 23*, 3 (2004), 309–314.

[RT00] Ruzon M. A., Tomasi C.: Alpha estimation in natural images. In *CVPR 2000* (June 2000), vol. 1, pp. 18–25.

[SB96] Smith A. R., Blinn J. F.: Blue screen matting. In *SIGGRAPH '96* (1996), ACM Press, pp. 259–268.

[Vid60] Vidor Z.: An infrared self-matting process. *Society of Motion Picture and Television Engineers*, 69 (June 1960), 425–427.

[Vla58] Vlahos P.: Composite photography utilizing sodium vapor illumination (u.s. patent 3,095,304), May 1958.

[Vla71] Vlahos P.: Electronic composite photography (u.s. patent 3,595,987, July 1971.

[Wal82] Wallace B.: Automated production techniques in cartoon animation, August 1982. Master's thesis, Cornell University.

[WBC*05] Wang J., Bhat P., Colburn R. A.,

Agrawala M., Cohen M. F.: Interactive video cutout. *ACM Trans. Graph. 24*, 3 (2005), 585–594.

[WC05] Wang J., Cohen M.: An iterative optimization approach for unified image segmentation and matting. In *ICCV* (2005), pp. 936–943.

[WGT*05] Wenger A., Gardner A., Tchou C., Unger J., Hawkins T., Debevec P.: Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph. 24*, 3 (2005), 756–764.

[WMPA97] Wolff L., Mancini T., Pouliquen P., Andreou A.: Liquid crystal polarization camera. *T-RA 13* (1997), 195–203.

[Wol94] Wolff L. B.: Polarization camera for computer vision with a beam splitter. *J. Optical Society of America A*, 11 (1994), 2935.

[YI02] Yahav G., Iddan G.: 3DV Systems' Zcam. *Broadcast Engineering* (June 2002).

[YNH04] Yasuda K., Naemura T., Harashima H.: Thermo-key: Human region segmentation from video. *IEEE Computer Graphics & Applications 24*, 1 (2004), 26–30.

[ZKU*04] Zitnick C. L., Kang S. B., Uyttendaele M., Winder S., Szeliski R.: High-quality video view interpolation using a layered representation. *ACM Trans. on Graphics 23*, 3 (2004), 600–608.

[ZWCS99] Zongker D. E., Werner D. M., Curless B., Salesin D. H.: Environment matting and compositing. In *SIGGRAPH '99* (1999), ACM Press/Addison-Wesley Publishing Co., pp. 205–214.