Color invariant chroma keying and color spill neutralization for dynamic scenes and cameras

Anselm Grundhöfer · Daniel Kurz · Sebastian Thiele · Oliver Bimber

Abstract In this article we show how temporal backdrops that alternately change their color rapidly at recording rate can aid chroma keying by transforming color spill into a neutral background illumination. Since the chosen colors sum up to white, the chromatic (color) spill component is neutralized when integrating over both backdrop states. The ability to separate both states additionally allows to compute high-quality alpha mattes. Besides the neutralization of color spill, our method is invariant to foreground colors and supports applications with real-time demands. In this article, we explain different realizations of temporal backdrops and describe how keying and color spill neutralization are carried out, how artifacts resulting from rapid motion can be reduced, and how our approach can be implemented to be compatible with common real-time post-production pipelines.

Keywords Chroma keying · Color spill · Image processing · Digitization and image capture

1 Introduction

Keying (or matting) is one of the most fundamental tools for image-, broadcast-, and film-production. It allows one to separate the foreground from the background in captured images or recorded video footage and to compose new images or video footage from the extracted foreground and a synthetic background. Many visual effects rely on high-quality keying. The result of the keying process is usually an alpha matte (α) that stores an opacity value for each pixel. The composite (C) is computed from a linear combination of foreground (F) and background (B), such as $C = \alpha F + (1 - \alpha)B$ [21]. A large variety of keying techniques exist, such as luma and chroma keying, difference and depth keying, polarization and defocus keying, and natural image matting techniques. All of them have distinct advantages and limitations. They will be discussed and compared with our method in Sect. 4 at the end of this article.

Today, chroma keying [26] is the most common method for generating professional alpha mattes by using colored (blue or green) backdrops. The backdrops are usually painted walls or colored curtains. Its dependence on foreground color (foreground objects with a color similar to the background cannot be keyed) and problems of color spill (i.e., visible color spreading from the green or blue background onto the foreground in the final composite) are the two main disadvantages of chroma keying.

Contribution We present an enhanced chroma keying method that is based on the fast temporal switching of backdrop colors. Compared to standard chroma keying, disturbing color spill is transformed into a neutral white background illumination. In contrast to digital color spill suppression, our method is automatic, supports real-time rates, and reproduces all colors correctly. It is not variant to distinct optical modulation of foreground materials, such as

---

A. Grundhöfer (✉) · D. Kurz · S. Thiele
Bauhaus University Weimar, Bauhausstr. 11, 99423 Weimar, Germany
e-mail: grundhoe@uni-weimar.de

D. Kurz
e-mail: daniel.kurz@uni-weimar.de

S. Thiele
e-mail: sebastian.thiele@uni-weimar.de

O. Bimber
JKU Institute of Computer Graphics, Altenberger Straße 69, 4040 Linz, Austria
e-mail: oliver.bimber@jku.at

polarization, retro-reflection and infrared reflection. As opposed to conventional chroma keying, it is invariant to foreground colors, enables ad-hoc adjustments of scene lighting during recording, and facilitates uniform background illumination. Our approach supports dynamic foregrounds and freely moving cameras, is fully automatic, compatible to common production pipelines, and can be combined with camera motion-tracking techniques. This combination of features is not provided by any other related technique described above.

The remainder of this article is organized as follows: Sect. 2 gives a preliminary overview over the proposed method. Our approach of temporal chroma keying will be explained in detail in Sect. 2.1. A color seam correction required to eliminate errors will be presented in Sect. 2.2. In Sect. 2.3, different realizations of temporal backdrops are described. Next, a practical implementation of the complete method is given in Sect. 2.4. Results, limitations as well as a comparison to other color spill suppression methods are shown in Sect. 3. In Sect. 4 the related work will be discussed in detail. Finally, our approach is summarized and directions of future research are given in Sect. 5.

2 Color invariant chroma keying and color spill neutralization

Video frames which are recorded at a speed of, for instance, 30 fps can be exposed for up to a 1/30th of a second. During that time, captured reflectance and color spill are blended. If chroma keying is applied, the latter is less problematic for extremely fast camera or foreground motion, since it is mixed with traces of motion blur. Under normal conditions, however, it is clearly visible. Instead of integrating each frame at normal speed, we propose to halve integration time and, consequently, double the frame rate. Therefore, an adjustable, temporal backdrop illumination sequentially displaying orthogonal colors in the RGB color space has to be synchronized to the camera. Integrating two subsequently recorded sub-frames (C_a and C_b) digitally by averaging them (i.e., $C_{ab} = (C_a + C_b)/2$) to return to the original frame rate neutralizes the chromatic (color) spill component. The reason for this is, that adding two saturated and orthogonal backdrop colors will result in a homogeneous illumination spectrum when averaged. This mimics a neutral backdrop illumination which also neutralizes color spill in the foreground.

Contrary to the classical triangulation approach presented in [23], each sub-frame is chroma keyed individually and the maximum of both alpha mattes is used. This process is illustrated in Fig. 1, and examples are shown in Fig. 2. Using the maximum of both mattes makes this approach insensitive to unregistered image pairs. The method of Smith and Blinn [23] is discussed in detail in Sect. 4. It fails in this situation, as shown in Fig. 3, since it supports only static scenes and cameras. Instead of analyzing and suppressing spill in each frame like classical techniques [7, 26], our method neutralizes color spill primarily by integrating over two complementary spill contributions during recording. The result is equivalent to a neutral white background illumination. This requires temporal backdrops that can be switched at the speed of the camera's frame rate as explained in Sect. 2.3.

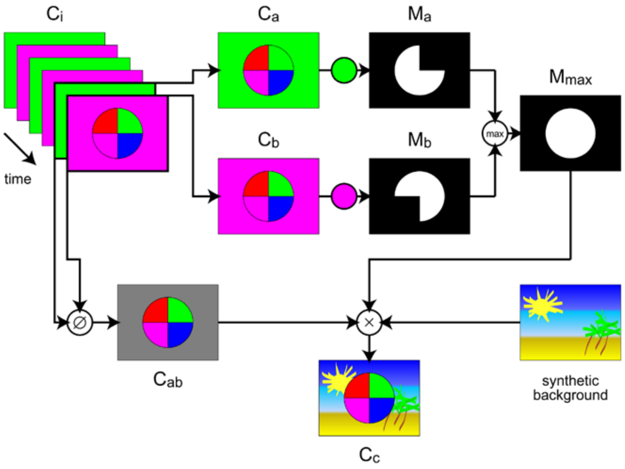


Fig. 1 Our approach: Sequentially recorded video frames C_i are separated into (C_a,C_b)-pairs according to the synchronously changing backdrop color. Chroma keying is applied independently to both frame sequences. The resulting alpha mattes M_a and M_b contain partially misclassified pixels when foreground and background colors match. The maximum ($M_{\max}$) of M_a and M_b is free of misclassifications. By blending the input frame pair into $C_{ab} = (C_a + C_b)/2$, color spill is transformed into a neutral background illumination. The final composite C_c is generated by applying $M_{\max}$ to C_{ab} and then combining the result with a synthetic background

2.1 Temporal chroma keying

As explained above, we pull an individual alpha matte (M_a and M_b) from each subframe (C_a and C_b). If the foreground color matches one of the background colors, then the corresponding matte will contain misclassified pixels. The classification of these pixels in the other matte, however, is always correct. Computing the maximum $M_{\max} = \max(M_a, M_b)$ results in an alpha matte that does not contain misclassified pixels. This matte is then used to compute the final composite image C_c, as displayed in Fig. 1 and demonstrated in Fig. 2.

Thus, our technique also becomes invariant to foreground colors, which is another relevant advantage over most classical chroma keying methods. This approach works well for static scene setups. However, if the content in C_a and C_b becomes misregistered during fast movements of the camera or the foreground, miscolored seam regions appear in C_{ab} and consequently also artifacts in C_c. This is displayed

Fig. 2 Chroma keying and blending the recorded image pair transforms color spill into a neutral background illumination (**c**) when complementary colors are used as backdrops (**a**), (**b**). The alpha mattes of (**a**) and (**b**) are generated individually (resulting in (**d**) and (**e**)) and are then combined to compute an optimized alpha matte (**f**). This neutralizes color spill and is independent of the foreground colors (compare (**g**) and (**h**) with (**i**)). Possible color artifacts during motion (close-up in (**c**)) are also corrected (close-up in (**i**)). Another example is shown in (**j–l**). Note the completeness of the alpha matte as well as the neutralized color spill in (**l**). Please zoom into individual image sections to see details of spill (*blue arrows*), keying (*yellow arrows*)

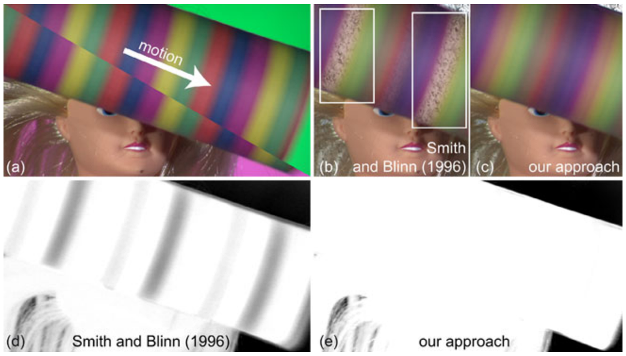


Fig. 3 The method proposed in [23] generates holes (**d**) and consequently imperfect composites (**b**) that result from misregistration during movement (**a**). These regions are classified correctly by our approach (**c**), (**e**)

in Figs. 4a–c, and shown in Fig. 2c. An additional post-processing step identifies these regions and color corrects the artifacts by reconstructing the missing foreground information as shown in Fig. 2i.

2.2 Color seam correction

Figure 4 outlines our color seam correction approach: First, we binarize M_{max} to M_{maxb} (cf. Fig. 4e). Then we find the contours in M_{maxb} by applying a Canny filter. Next we extrude these by dilation, as shown in M_{cext} in Fig. 4f.

By capturing with 59.94 Hz, we assume that the maximum motion in two consecutive frames is less than the dilated region in M_{cext}. Since we are only interested in the contour neighborhood of M_{maxb}, the intersection between M_{cext} and M_{maxb} gives us the regions with potential color seams, as shown in Fig. 4g. For all pixels within this intersection, we need to determine if their seam colors are caused by the background in C_a or by the background in C_b, or if they belong completely to the foreground (pixel 4 in Fig. 4i) or the background (pixel 1 in Fig. 4i). For the latter two cases, no additional actions are necessary, whereas the first two cases require special treatment. These situations can be determined by comparing the absolute RGB color difference between C_a C_b and the background colors B_a B_b:

$$s = \frac{|C_a - B_a| \cdot |C_a - B_b|}{|B_a B_b|} - \frac{|C_b - B_b| \cdot |C_b - B_a|}{|B_a B_b|}. \qquad (1)$$

Both B_a and B_b must be known in the camera's color space to normalize Eq. 1 with respect to the camera's response. They are determined by first applying the inverse of M_{maxb} to C_a and to C_b (to mask all background pixels) and then computing the two medians for the two sets.

If $|s|$ is above a threshold (T_s), then the corresponding pixel definitely belongs to a seam region. Since Eq. 1 is normalized, the threshold is constant for all frames. Empirically, we found that $T_s = 0.05$ is suitable for a wide variety of recording scenarios. If s is negative, the seam color is caused by the background recorded in C_a, and C_b contains the correct foreground pixel. If s is positive, the seam color is caused by the background recorded in C_b, and C_a contains the correct foreground pixel. Otherwise, if $|s|$ is below T_s, then C_a and C_b contain foreground pixels.

All pixels classified as being located in a color seam region are marked in a binary error matte M_{err} (cf. Fig. 4h). To avoid misclassifications of moving highlights, regions with saturated intensity values in C_a or C_b are not taken into account. We apply a morphological opening to M_{err} to clean voids that are caused by camera noise.

Simply replacing the color seam pixels in C_{ab} by the correct foreground pixel from C_a or C_b will not neutralize their

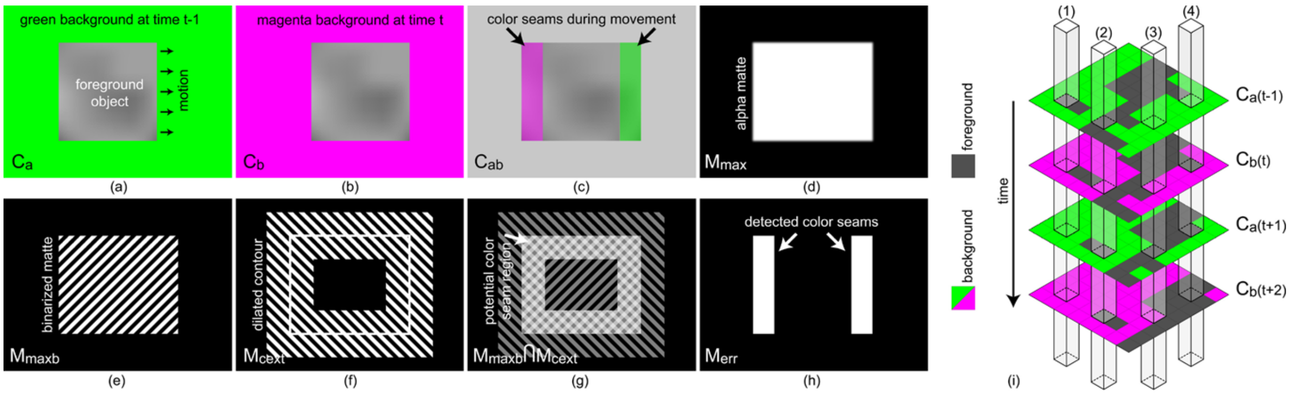


Fig. 4 Steps of the color seam correction as explained in Sect. 2.2: binarization (**e**), contour detection and dilation of contour (**f**), separation of color seam regions at the difference between $M_{\max}$ and M_{cext} (**g** and **h**). In (**i**), the four possibilities of pixel processing for times $t-1$ through $t+2$ are illustrated. For pixels 1 and 4, M_{err} does not contain an entry. Thus, their intensities are blended without any further processing. For pixel 2, $C_b(t)$ contains background information. Therefore, $C_b(t+2)$ is analyzed in a first step. In this example, it contains foreground information and can therefore be blended with $C_a(t+1)$. For pixel 3, however, $C_b(t)$ as well as $C_b(t+2)$ contain background. Thus, blending would not avoid the color seams in this case. Instead, the spatial neighborhood of pixel 3 in $C_b(t)$ and $C_a(t)$ is analyzed and the neighbor with the most similar intensity is used to compensate the color seam as described in Sect. 2.2

color spill. We correct this by processing only pixels that are marked in M_{err} as follows (cf. Fig. 4i):

If a pixel in C_b was captured at time t and contains background, and the corresponding pixel in C_a was captured at $t+1$ and contains foreground (as illustrated in Fig. 4i) then we compare the foreground pixel in C_a at $t+1$ with the corresponding pixel in C_b at $t+2$, using Eq. 1. If the latter is classified as foreground, both pixels are averaged to neutralize their color spill and the result is stored in C_{ab}. The same procedure is carried out with a pixel in C_a at $t-1$, if the corresponding pixel in C_b at t contains foreground and C_a at $t+1$ contains background. In this case, the pixel intensities in C_a at $t-1$ and C_b at t are averaged.

In the rare case of quickly moving thin foreground objects, this procedure is invalid, because a look back or a look ahead in the frame sequence will point to a background instead of foreground pixel. Again, this situation can be determined using Eq. 1. If this applies, we perform a spirally increasing spatial search around the original foreground pixel within the subset marked in M_{err} to find the closest neighbor that Eq. 1 identifies as a foreground pixel in both C_a and C_b. We do not simply average them as before to neutralize color spill, but try to preserve most of the information in the original pixel. Instead, we assume that the foreground color of the original (at location x, y) and the selected neighbor (at location i, j) may vary slightly, but their color spills are identical. We extract the color spill of the neighbor with $C_s(i,j) = (C_a(i,j) + C_b(i,j))/2 - C_b(i,j)$ if $C_a(x,y)$ contains the original foreground pixel, or with $C_s(i,j) = (C_a(i,j) + C_b(i,j))/2 - C_a(i,j)$ if $C_b(x,y)$ contains the original foreground pixel. The color spill of this pixel can then be suppressed by $C_{ab}(x,y) = C_a(x,y) - C_s(i,j)$ or $C_{ab}(x,y) = C_b(x,y) - C_s(i,j)$ – depending on whether the original foreground was in C_a or in C_b.

Since the capturing rate is twice the final video rate, looking ahead by one captured frame leads to a constant delay of 1/2 video frame. If this delay is not tolerable (e.g., for live broadcasting), then averaging with a corresponding foreground pixel in the look-ahead frame can be replaced by the spiral spatial search in the present frame, as explained above for quickly moving thin foreground objects. To compensate for the required larger number of search steps, the spatial search region can be reduced.

2.3 Temporal backdrops

The required synchronized switching of the backdrop at a high speed can be realized in various ways: Video projectors with a low latency, such as DLP projectors, can be used in a front- or a rear-projection setup. To avoid temporal artifacts that result from a color wheel, 3-chip DLP projectors should be preferred over single chip projectors. Projector pairs with higher latencies are also applicable in combination with fast LCD shutters in front of their lenses. Applying projectors has the advantage that temporal backdrops can be generated on non-trivial scene geometry, as commonly used in studios. Additionally, low-latency flat-panel displays can be employed. Commercial LCD, Plasma and FED panels support frame rates of up to 240 Hz, while the refresh rate of upcoming OLED and PLED displays might be even higher because of their significantly lower switching time of <1 µs. Temporal backdrops that are realized with a display technology can be coded temporally and spatially. The additional spatial coding can support camera tracking as outlined in Sect. 5. However, if this option is not required,

Fig. 5 Different prototype implementations of temporal backdrops: a backprojection screen (**a**), a softbox with integrated RGB LEDs (**b**), and a RGB front-illumination of a diffuse background surface (**c**). The foregrounds are lit with arbitrarily aligned spot lights

a simple temporal illumination of the background represents a cost-efficient alternative. This can be realized with RGB LEDs that illuminate white background diffusors, similar to [29], yet with alternating colors. Softboxes in the background that employ RGB LEDs for illumination is a further option. Another possible solution would be using reflective colored electronic ink displays. Although these devices currently suffer from very high latencies, faster devices may become available in the near future[1] [30]. In this case, the background has to be lit homogeneously, much like a standard blue screen.

No matter which approach is chosen, the backdrop has to be synchronized with the shutter of the camera and the complementary colors being used have to add up to white. Figure 5 shows several of our current implementations of temporal backdrops.

2.4 Implementation

To synchronize the temporal backdrops with the shutter of the camera (a high definition 3CCD JVC GY-HD251 in our case), we utilize the composite output signal of the camera as an external trigger. If the backdrop is realized with displays (e.g., projectors or flatpanel screens), then we currently apply a Quadro FX 4500 X2 G-Sync whose G-sync input is connected with the composite output of the camera. The graphics card renders the images B_a and B_b. To synchronize LED-driven backdrops, we use a custom-built electronic controller that is triggered by the VGA-out signal of the graphics card. All 720p frames are captured to disk via HD-SDI in Yuv 4:2:2 format at HD scanning speed of 59.94 Hz. Cameras with Bayer pattern and interlaced cameras require an interpolation in color channels or scanlines respectively to ensure an equal resolution in C_a and C_b. Because of the required precise synchronization of the camera and the temporal backdrops, systems with rolling shutters cannot be applied. Cameras with 3 CCD sensors and global shutter that record progressive images are applied in our implementation, since they provide the highest image quality.

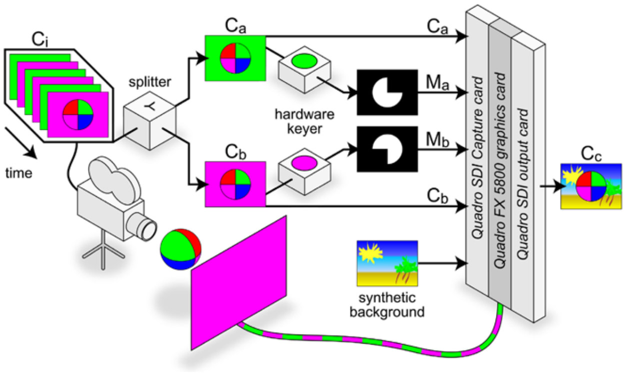


Fig. 6 Studio-compatible hardware implementation of temporal backdrops: a splitter separates both backdrop states and forwards the corresponding images to hardware keying units. The computed alpha mattes and the original frames are forwarded to an Nvidia Quadro digital video pipeline that executes our color seam correction algorithm in real-time

We chose green (G) and magenta (RB) as backdrop colors for the reason described in Sect. 3, and we apply Ultimatte software chroma keying to compute M_a and M_b. All other processing steps explained in Sect. 2.2 are implemented as Cg fragment shaders and are carried out in an average processing time of less than 17 ms (without texture up- and downloads) for high definition video resolution on an Nvidia Geforce 285 GTX. In our proof-of-concept implementation, we pre-record the video frames and apply software keying and compositing, since these steps are independent of our algorithm.

Figure 6 sketches a possible solution for integrating our technique easily into existing professional hardware composition pipelines to support real-time keying. The recorded video signal has to be split into C_a and C_b sequences using, for example, a Blackmagic Design Multibridge Pro module. The two sequences can then be streamed to two hardware keying units, such as an Ultimatte 11 HD/SD. Our algorithm can be executed at the pipeline's overall clock speed on an additional PC that is equipped with a genlocked Nvidia Quadro digital video pipeline system: The sequences of alpha mattes M_a and M_b, and the C_a and C_b sequences can

---

[1] http://www.pixelqi.com/products.

be forwarded from the keying units to a Quadro SDI Capture card for streaming them directly onto the GPU memory of one or multiple Quadro FX 5800 graphics cards that are used for image processing and compositing. Additional Quadro graphics boards can be employed to drive all backdrop displays in sync with all other components. The final composite sequence can then be read out via a Quadro SDI Output card.

Alternatively, the entire pipeline (including keying, seam correction, spill neutralization and compositing) can be implemented on the GPU, if compatibility to established professional soft- and hardware components is not required.

3 Results and limitations

In Fig. 7, we compare our color spill neutralization to a uniformly illuminated white backdrop as ground truth. We also determine the color spill generated by single green and magenta backdrops, and evaluate how well it can be reduced by professional spill suppression algorithms (Ultimatte Advant Edge, library version 1.6.2 in our case). Table 1 provides the measured ΔE^*_{ab} chrominance differences to the ground truth for the samples shown in Fig. 7. Our method clearly outperforms professional spill suppression techniques.

Table 1 ΔE^*_{ab} chromacity differences for the sample shown in Fig. 7

Input	Average ΔE^*_{ab}	Maximum ΔE^*_{ab}
Green	11.3	26.7
Magenta	11.7	27.7
Green ultimatte	7.0	32.8
Magenta ultimatte	9.3	50.4
Proposed method	2.6	5.0

While, on average, such techniques reduced the ΔE^*_{ab} of visible color spill only by a factor of 1.4 in our experiments, our technique compensated its ΔE^*_{ab} by a factor of approximately 4.4—down to $\Delta E^*_{ab} = 2.6$. Note that a ΔE^*_{ab} of 2.3 represents the visibility threshold and equals one just noticeable difference (JND) [22]. Note that the software spill suppression used in Ultimatte fails for foreground pixels with color values similar to the backdrop, as shown in Fig. 7. This is not the case when our approach is applied. As explained above, we transform color spill into a neutral (white) background illumination by averaging C_a and C_b. It is also imaginable to entirely remove spill by subtracting the absolute values of $C_a - C_b$ from C_{ab}. This would mimic a neutral black background, but is only possible for entirely static setups, since non-registered pixel intensities with different colors will lead to a false color reproduction.

It was shown in [10] as well as in [29] that applying motion interpolation for computing intermediate frames is efficient for time-multiplexed techniques with no real-time demands. For post-processing applications, this allows to further reduce the color seam regions in our case.

Additionally, chroma keying can be combined with a natural feature matting, such as Bayesian image matting [5]. This can be fully automated by applying an eroded and a dilated version of M_{maxb} to compute a trimap, as illustrated in Fig. 8.

In all our prototype setups, the backdrops are temporally switched at 59.94 Hz. Although this is above the critical flicker frequency of the human visual system [19], slight alternating backdrop intensities can be perceived. This is reduced to a minimum by choosing green and magenta as backdrop colors, since their integrated luminance can be matched easily as explained in [32]. The remaining brightness alternations are barely detectable. Color seams are observable at the edges of the backdrops during fast eye move-

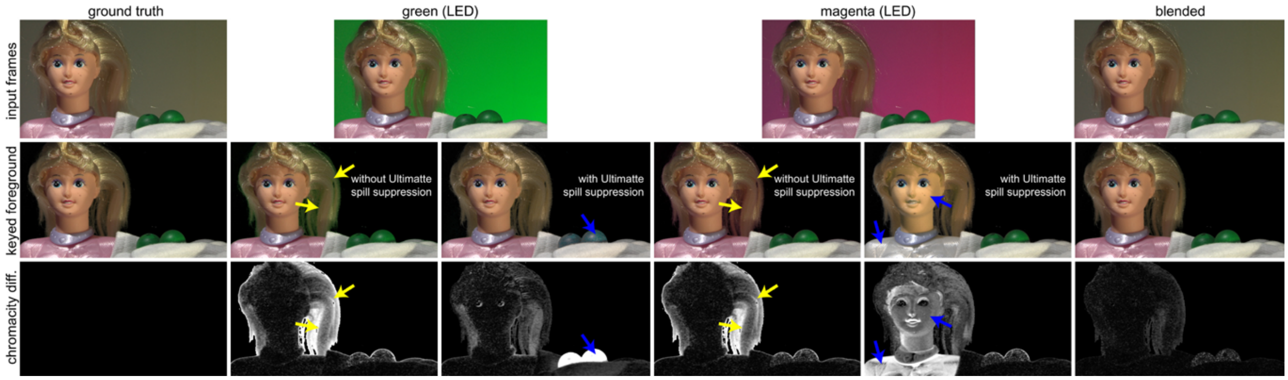


Fig. 7 Comparison of color spill neutralization and spill suppression: a scene with a uniformly lit white background serves as ground truth (*left column*). The remaining color spill is calculated from the ΔE^*_{ab} chrominance distance in CIE L*a*b* color space—with and without software spill suppression (*two center columns* for *green and magenta backgrounds*). The result of our color spill neutralization is shown in the right column. Please zoom into individual image sections to see details. The *yellow arrows* indicate errors caused by remaining color spill, while the *blue arrows* point to wrong color reproductions caused by spill suppression

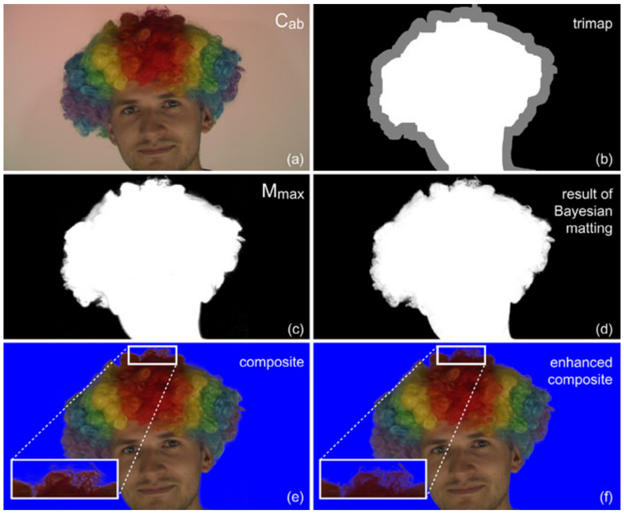


Fig. 8 Keying enhancements through additional Bayesian matting for applications without real-time demands: A trimap (**b**) can be computed automatically by eroding and dilating M_{maxb}. After applying Bayesian matting to C_a and C_b, the maxima of both results are used to calculate improved alpha mattes (**d** and **f**) as does chroma keying alone (**c** and **e**). Please zoom into individual image sections to see details. Note the reduced white seam resulting from wrong alpha values in the close-up

ment. Both, however, do not distract people when working with temporal backdrops.

The irradiance of the backdrop is mixed with the illumination of the foreground. This leads to a reduced color saturation that depends on the amount of light arriving at the backdrop. In our experiments, however, we measured a foreground illumination arriving at the backdrop that was more than twice the actual backdrop brightness (110 Lux vs. 48 Lux) but did not affect the quality of the resulting alpha mattes.

Nevertheless, our method has limitations: One fundamental drawback of chroma keying is that strongly reflecting or refracting foreground objects are not keyed correctly. Furthermore, in the rare case that moving foreground objects contain neighboring complementary colors that are identical to our backdrop colors, keying also fails. Strong motion blur resulting from rapid movements can lead to lower keying quality. This, however, is also a common problem for standard chroma keying with static backdrops. High-quality motion interpolation and shorter shutter times will reduce this effect.

While geometrically complex studio props can, in principle, be transformed into temporal backdrops using front- or back-projection, they are easier to manufacture from uniformly colored material that can be keyed. Such physical props are often used to give actors or moderators orientation with respect to virtual props. This, however, can alternatively be achieved with a synchronized projection, as explained in [10]. For outdoor sets, static backdrops might still be more efficient than temporal backdrops.

The following section provides a detailed summary and comparison of related work focusing on keying techniques that are applied in controlled environments, such as studios.

4 Related work

Several techniques exist that neither require controlled illumination nor controlled backgrounds. Examples are the application of time-of-flight sensors [12] that results in low quality binary mattes. Semi-automatic natural image matting techniques [5, 24] require user interaction for defining trimaps and, hence, do not support real-time keying. Automatic natural image matting techniques apply additional constrains, such as defocus [16], motion [18] or depth [28] information for estimating trimaps without user interaction. These techniques, however, are restricted to the presence of visually distinguishable features in the foreground and background, and suffer from the same limitation as semi-automatic methods in terms of similar foreground and background colors. These techniques are out of scope of this article, since high-quality results can be achieved easier by controlling backgrounds or illumination in the context of studio environments. Furthermore, color spill suppression is a technique to reduce undesirable color spill. It was first addressed in [26], where a solution is presented for blue screen matting by suppressing the blue color channel in the foreground pixels. This can lead to wrong color reproduction of the foreground in the final composite. Information about the background and lighting color is required to estimate and remove the amount of color spill by a least-squares solution in [7]. However, since manual user interaction is needed to classify regions with possible color spill, this method cannot be used in real-time. More advanced spill suppression algorithms were presented in [27] and [2], but both require capturing the empty background for each new recording perspective and therefore only support a static camera and fixed lighting conditions. The subsequent sections categorizes related work that apply either controlled backgrounds or a controlled illumination for matting.

Controlled backgrounds In [23], two different backdrops and triangulation matting are used to calculate alpha mattes for an arbitrarily colored foreground. Originally, this method is not suitable for dynamic video recording, because it requires re-capturing the two empty backgrounds without foreground every time the camera is moved. Defining constant backdrop colors in advance for extending this algorithm to support camera movements is imaginable. Moving textured foreground, however, will still be classified incorrectly if the corresponding pixels in both images contain different color values. This is explained in Sect. 2 and demonstrated in Fig. 3. Furthermore, color spill cannot be handled. An enhancement of this method is presented in [14], where color ramps with different frequencies are displayed as backdrops to improve the quality of the generated alpha matte in the case of highly specular materials. This method, however, requires 12 images for each alpha matte to be recorded—6 with the background ramps and 6 with the foreground objects added. Thus, the approach is not applicable for real-time recording, but was used in the context of offline 3D reconstruction. The neutralization of color spill in video compositions is not addressed. Environment matting [32] is an extension of alpha matting that models reflection (including spill) and refraction of light from the backdrop plane with the foreground correctly. For calibration, most variants of environment matting apply single or multiple backdrop screens and display Gray codes or color ramps [3, 20, 31]. Since many images must be captured, dynamic foregrounds and camera movements cannot be supported. A simplification with one image has been described in [4] which makes environment matting applicable for dynamic situations within completely controlled illumination environments, but with this method color spill cannot be handled at all. In [13], a time-multiplexed (on/off) background was also applied to support simple threshold-based keying. Besides relatively poor results, this technique is variant to low-reflectance foregrounds and does not address spill. In [29], a similar approach is followed. A variation of defocus keying was presented in [17], that applies a backdrop containing high frequent spatial color patterns that are sequentially captured with a wide (out of focus) and a narrow (in focus) aperture. Since the foreground is always in focus, an alpha matte can be computed from the per-pixel color difference between both images. The camera movement is limited in this approach, since the background has to remain in focus during the narrow aperture state. Spill is not addressed, and a special camera with a high-speed aperture is required.

Controlled illumination Other methods apply polarizing [13, 15, 17] or polarization-preserving [1] backdrops instead of colored ones. Two synchronized and optically aligned cameras with differently rotated polarization filters are used in these cases. While the filters restrict the cameras' orientation along the optical axis to remain aligned with the polarization direction of the backdrop, homography-based camera registration adds additional limitations to varying foreground depth and camera distances. For polarization-preserving backdrops, the entire illumination must be polarized. Color spill is generally not an issue for polarization keying. But like chroma keying, it suffers from variance of foreground polarization and spill of polarized light from the background onto a polarization preserving foreground. Instead of polarization, retro-reflection can be utilized. Retro-reflective backdrops reflect light in a narrow solid angle back to its source. Lighting such backdrops with blue or green light directly from the camera (e.g., using an LED ring) ensures a homogeneously colored background [9]. Compared to diffuse screens, color spill from the background is reduced due to the highly directed retro-reflection. The foreground, however, is directly lit by the colored LED ring, which results in similar problems as that of color spill which is indirectly reflected from diffuse backdrop. A careful foreground illumination is necessary to overcast it. Infrared light (IR) can also be applied to illuminate the backdrop while the foreground is lit by white light. An optically aligned IR and RGB color camera pair then allows pulling mattes from the invisible IR spectrum. One implementation of this was presented in [6] for static cameras. Compared to chroma keying, color spill and color variance are not problematic. But instead, foregrounds with high IR reflectance, such as several clothes, can cause IR spill that influences keying. Synchronized video projectors and LED illumination are used in [10] to pull mattes in front of an arbitrary background. In every other frame, a compensation image is projected onto background surfaces to neutralize their natural reflectance while the foreground illumination is turned off. Luma keying is applied, and the composite can be computed from the remaining frames for which the foreground and background are illuminated while the projection is turned off. This method is independent of foreground colors, but does not consider color spill. The need for front-projection, which leads potentially to shadows of foreground objects on the background, makes this method less suitable for production sets with space constraints. Furthermore, the entire foreground illumination as well as emissive foreground objects (such as monitors) must be controllable and have to be synchronized with the recording. Misregistrations that result from fast movements during two subframes lead to inconsistent alpha mattes if no additional motion interpolation is applied. This, however, cannot be realized in real-time. In addition to neutralizing color spill, our approach also overcomes all of these problems. Using a flash-on/flash-off image pair for matting was proposed in [25]. This technique is limited to binary mattes that are constrained to opaque foreground objects. Camera-synchronized blue LEDs are used in [8] to illuminate the foreground in every other frame in addition to the regular scene lighting. Thus, chroma keying can be applied to pull alpha mattes from the blue-colored foreground that is recorded in every other frame. This method is vulnerable to foreground colors. Synchronized white LEDs are used in [11] to flash the foreground in every other frame. Since difference keying is applied, this approach is unreliable for low-reflecting foregrounds.

5 Summary and future work

We have presented a novel matting approach that improves conventional chroma keying in two ways: firstly, our color spill neutralization method leads to significantly better color reproduction than digital color spill suppression techniques, is fully automatic and supports real-time rates. Secondly, the temporally switched complementary colors make chroma keying invariant to any foreground color, and a uniform and consistent background illumination is easier to achieve with emissive or rear-illuminated backdrops than with reflective ones that are front-illuminated. It allows an ad-hoc adjustments of scene lighting while remaining robust keying results.

In contrast to related approaches that are alternative to standard chroma keying, our method is not variant to distinct optical modulation of foreground materials, such as polarization, retro-refection and infrared reflection, supports dynamic foregrounds and camera motion. It is fully compatible with conventional hardware composition pipelines which are used in professional studios, and can be combined with motion-tracking techniques.

The latter is demonstrated in Fig. 9, where a display is used as temporal backdrop for spatially encoding static feature points in addition to a temporal color coding for supporting simultaneous camera motion tracking. The feature points are displayed during both backdrop states and have the same chrominance as the corresponding backdrop color, but are slightly reduced in luminance. As explained in [10], they can then be used as input for classical matchmoving algorithms that reconstruct the 3D camera path or for motion-tracking techniques that determine 2D displacement of the camera with respect to the background plane. In our experiments, a luminance reduction of 15% was sufficient for robust feature detection without affecting the matting process.

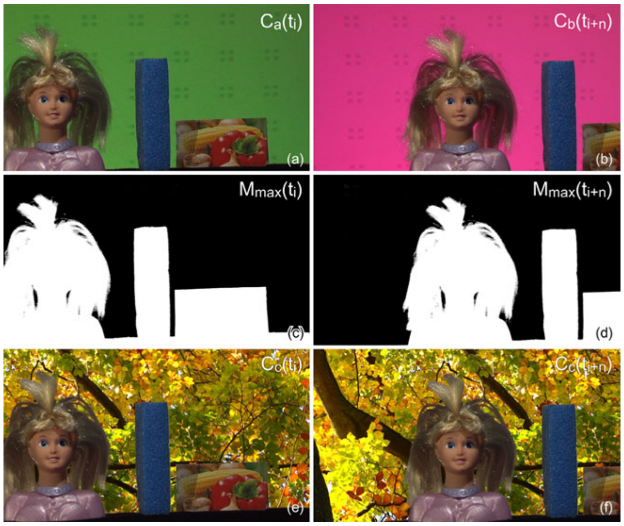


Fig. 9 Supporting camera motion tracking through embedded spatial codes: frames C_a at time t_i (**a**) and C_b at time t_{i+n} (**b**) with embedded feature points. Reducing the luminance of the displayed code features does not influence the quality of the computed alpha mattes (**c** and **d**). The final composites (**e** and **f**) display the result of a perspectively augmented synthetic background as a result of motion tracking

Besides professional visual effects production, we envision applications for temporal backdrops in other domains such as novel game interfaces. For instance, video cameras such as Sony's EyeToy are becoming ever more popular components of gaming consoles to integrate players directly into games. A simple LED front illumination that temporally flashes the walls behind the player (similar to that illustrated in Fig. 5c at a smaller scale) might help to make the keying process more robust. Visually, this is perceived as white illumination.

Acknowledgements This project is supported by the Deutsche Forschungsgemeinschaft (DFG) under reference number BI 835/2-1.

References

1. Ben Ezra, M.: Segmentation with invisible keying signal. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, vol. 1, pp. 32–37. IEEE Computer Society, Los Alamitos (2000)
2. Chaplin, D.J.: Chroma keyer with fringe control offset. uS Patent number 5,313,304 (1994)
3. Choudhury, B., Singla, D., Chandran, S.: Fast color-space decomposition based environment matting. In: I3D '08: Proceedings of the 2008 Symposium on Interactive 3D Graphics and Games, p. 1. ACM, New York (2008)
4. Chuang, Y.Y., Zongker, DE, Hindorff, J., Curless, B., Salesin, D.H., Szeliski, R.: Environment matting extensions: towards higher accuracy and real-time capture. In: SIGGRAPH '00: ACM SIGGRAPH 2000 Papers, pp. 121–130. ACM, New York (2000)
5. Chuang, Y.Y., Curless, B., Salesin, D.H., Szeliski, R.: A Bayesian approach to digital matting. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR, vol. 2, pp. 264–271. IEEE Computer Society, Los Alamitos (2001)
6. Debevec, P., Wenger, A., Tchou, C., Gardner, A., Waese, J., Hawkins, T.: A lighting reproduction approach to live-action compositing. ACM Trans. Graph. **21**(3), 547–556 (2002)
7. Dupont, J., Deschenes, F.: Toward a realistic interpretation of blue-spill for blue-screen matting. In: CRV '06: Proceedings of the 3rd Canadian Conference on Computer and Robot Vision, p. 33. IEEE Computer Society, Washington (2006)
8. Graham, T., Richard, R.: Flash-based keying. European Patent Application EP1499117 (2005)
9. Grau, O., Pullen, T., Thomas, G.: A combined studio production system for 3-d capturing of live action and immersive actor feedback. IEEE Trans. Circuits Syst. Video Technol. **14**(3), 370–380 (2004)
10. Grundhöfer, A., Bimber, O.: Virtualstudio2go: digital video composition for real environments. In: SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 Papers, pp. 1–8. ACM, New York (2008)
11. Grundhöfer, A., Seeger, M., Häntsch, F., Bimber, O.: Dynamic adaptation of projected imperceptible codes. In: ISMAR '07: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 1–10. IEEE Computer Society, Washington (2007)
12. Gvili, R., Kaplan, A., Ofek, E., Yahav, G.: Depth keying. In: Woods, A.J., Bolas, M.T., Merritt, J.O., Benton, S.A. (eds.) Stereoscopic Displays and Virtual Reality Systems X. SPIE, vol. 5006, pp. 564–574. SPIE, Bellingham (2003)
13. Kuechler, M., Kunz, A.: Imperceptible projection blanking for reliable segmentation within mixed reality applications. In: Kjems,

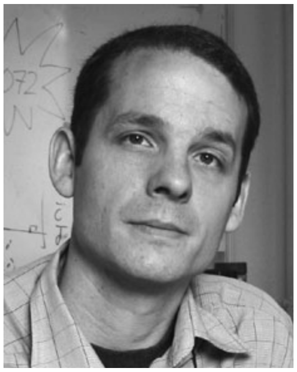
E., Blach, R. (eds.) 9th International Workshop on Immersive Projection Technology, 11th Eurographics Workshop on Virtual Environments, pp. 23–30. Eurographics Association, Denmark (2005)

14. Matusik, W., Pfister, H., Ngan, A., Beardsley, P., Ziegler, R., McMillan, L.: Image-based 3d photography using opacity hulls. ACM Trans. Graph. **21**(3), 427–437 (2002)
15. McGuire, M., Matusik, W.: Real-time triangulation matting using passive polarization. In: SIGGRAPH ’06: ACM SIGGRAPH 2006 Sketches, p. 88. ACM, New York (2006)
16. McGuire, M., Matusik, W., Pfister, H., Hughes, J.F., Durand, F.: Defocus video matting. In: SIGGRAPH ’05: ACM SIGGRAPH 2005 Papers, pp. 567–576. ACM, New York (2005)
17. McGuire, M., Matusik, W., Yerazunis, W.: Practical, Real-Time Studio Matting Using Dual Imagers, pp. 235–244 (2006)
18. Moon, J.H., Kim, D.O., Park, R.H.: Video matting based on background estimation. In: World Academy of Science, Engineering and Technology, vol. 2 (2005)
19. RA, Moses, Hart, W.M.: The temporal responsiveness of vision. Adler’s Physiology of the Eye (1987)
20. Peers, P., Dutré, P.: Wavelet environment matting. In: EGRW ’03: Proceedings of the 14th Eurographics Workshop on Rendering, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, pp. 157–166 (2003)
21. Porter, T., Duff, T.: Compositing digital images. In: SIGGRAPH ’84: ACM SIGGRAPH 1984 Papers, pp. 253–259. ACM, New York (1984)
22. Sharma, G.: Digital Color Imaging Handbook. CRC Press, Boca Raton (2002)
23. Smith, A.R., Blinn, J.F.: Blue screen matting. In: SIGGRAPH ’96: ACM SIGGRAPH 1996 Papers, pp. 259–268. ACM, New York (1996)
24. Sun, J., Jia, J., Keung Tang, C., Yeung Shum, H.: Poisson matting. In: SIGGRAPH ’04: ACM SIGGRAPH 2004 Papers, vol. 23, pp. 315–321. ACM, New York (2004)
25. Sun, J., Sun, J., Kang, S.B., Xu, Z.B., Tang, X., Shum, H.Y.: Flash cut: foreground extraction with flash and no-flash image pairs. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2007)
26. Vlahos, P.: Electronic composite photography with color control. uS Patent number 4,007,487 (1977)
27. Vlahos, P.: Backing color and luminance nonuniformity compensation for linear image compositing. uS Patent number 5,032,901 (1991)
28. Wang, O., Finger, J., Yang, Q., Davis, J., Yang, R.: Automatic natural video matting with depth. In: PG ’07: Proceedings of the 15th Pacific Conference on Computer Graphics and Applications, pp. 469–472. IEEE Computer Society, Washington (2007)
29. 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), 756–764 (2005)
30. Whitesides, T., Walls, M., Paolini, R., Sohn, S., Gates, H., McCreary, M., Jacobson, J.: 10.2: Towards video-rate microencapsulated dual-particle electrophoretic displays. SID Symp. Dig. Tech. Pap. **35**(1), 133–135 (2004)
31. Zhu, J., Yang, Y.H.: Frequency-based environment matting. In: PG ’04: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference, pp. 402–410. IEEE Computer Society, Washington (2004)
32. Zongker, D.E., Werner, D.M., Curless, B., Salesin, D.H.: Environment matting and compositing. In: SIGGRAPH ’99: ACM SIGGRAPH 1999 Papers, pp. 205–214. ACM, New York (1999)

Anselm Grundhöfer received the Diploma (M.Sc.) degree in media system sciences from Bauhaus-University Weimar, Germany, in 2006. Currently he is a Ph.D. candidate at Bauhaus-University Weimar. His research interests include projector–camera systems, real-time image processing, computer vision and display technologies.

Daniel Kurz graduated with a M.Sc. degree in media system sciences from the Bauhaus-University Weimar in early 2010. Currently he is working in the research and development group of metaio, Munich. His research interests include augmented reality, computer vision and computer graphics.

Sebastian Thiele received a B.Sc. in media systems science from Bauhaus-University Weimar in 2010. Currently he is a M.Sc. student at Bauhaus-University Weimar. His research interests include real-time rendering, scientific visualization, computer vision and display technologies.



Oliver Bimber has led the Augmented Reality Group at the Bauhaus-University Weimar until March 2010. Since October 2009, he is heading the Institute of Computer Graphics at the Johannes Kepler University Linz, Austria. More information: www.jku.at/cg