

Robust Image Chroma-Keying: A Quadmap Approach Based on Global Sampling and Local Affinity

Wenyi Wang and Jiying Zhao, *Member, IEEE*

Abstract—Chroma-keying is a technique used to replace solid-colored background of images or video frames. This technique is widely used in TV broadcasting, film production, augmented reality, and virtual environment. This paper proposes a new chroma-keying method, which can automatically remove the background color in an image and accurately segment the foreground objects along with their transparency property. Compared to conventional chroma-keying methods based on color clustering, color difference, or thresholding, the proposed method analyzes the color statistics and color confidence of the image in global range. By analyzing image color statistics, local lightness variation, and experience from human visual perception in HSV color space, a segmentation map called quadmap is automatically generated to segment the image into four types of regions: 1) foreground; 2) background; 3) transparent; and 4) reflective regions. By using quadmap, the proposed chroma-keying system can differentiate between transparent and reflective regions. This has always been a challenging problem in conventional chroma-keying or α matting systems. This improvement generates more reliable foreground colors in reflective regions. As a result, there can be less constraints for foreground scene used in TV-broadcasting or film making. The procedures of the proposed method consist of the following five steps: 1) background region detection based on color statistics and local lightness variation; 2) absolute foreground region detection based on the knowledge of background color and predefined thresholds in Hue channel of HSV color space; 3) reflective region detection based on experience from human visual system; 4) background color propagation based on Laplacian equation; and 5) transparency value and foreground color estimation based on global sampling and color confidence.

Index Terms—Chroma-keying, histogram analysis, global sampling, lightness variation, human visual system (HVS), Laplacian equation.

I. INTRODUCTION

IMAGE composition [1] is the technique which accurately and smoothly combines the foreground objects (such as actors/actresses and broadcasters) and the background scene that is taken somewhere else or generated by computer. This technique plays an important role in TV broadcasting, film production, augmented reality and virtual environment,

Manuscript received November 24, 2014; revised February 27, 2015; accepted March 2, 2015.

W. Wang and J. Zhao are with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada (e-mail: jzhao@uottawa.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBC.2015.2419181

because of its good performance, flexible capability of integrating real objects with computer generated scenes as well as the cost saving by avoiding outdoor filming.

Before the foreground is composited with desirable background scene, the foreground scene shot in the studio must be finely segmented, especially for reflective parts, transparent regions and boundaries with color spill. For still image editing, α matting algorithms have been extensively researched for decades [2], [3]. The capability of dealing with nature images makes α matting universal at the cost of high computational complexity and sometimes tedious manual operation. However, nature image matting is not the only concern in image/video editing because of the requirements of real-time processing, automaticity and user friendliness. With these concerns, chroma-keying is the most widely applied video segmentation technique in broadcasting, advertising and film making. In chroma-keying system, foreground is shot in front of solid color background (usually blue or green) with the concern of background color simplicity and foreground-background color distinction. Given the solid background video sequence and the background color, a chroma-keying system “keys” out the foreground objects along with their transparency property. This is done by removing all pixels with similar chromaticity and lightness as the provided background color and by estimating the transparency factor for pixels with color mixed with background color. Although the ideal background color is supposed to be solid, the actual background color changes along with studio environment such as light, shadow and rugged background. Furthermore, chroma-keying becomes even more challenging if transparent and reflective regions coexist in foreground scene.

The solutions for nature image matting are compelling in recent years. Many methods were proposed from different aspects such as sampling based [4]–[6], statistics based [7] and affinity based [8], [9]. The chroma-keying problem, however, does not draw as much attention as alpha matting does in recent years. Agata *et al.* [10] proposed to use a two-tone checker pattern background to overcome the problem that foreground objects become transparent if their colors are the same as the background color. Vidal [11] proposed to project two polarization-orthogonal images onto the retroreflective background screen to provide immersive feedback. These recent chroma-keying researches mainly focused on application issues instead of color analysis. Many researches on color analysis for chroma-keying can date back to decades

ago [12]–[14]. However, most chroma-keying systems used today are still fundamentally based on one or more early proposals. Luma-keying [15] is a fast and simple chroma-keying method, which extracts foreground objects with respect to a predefined luminance threshold. The color blindness of this method makes it not suitable in many situations except for a few applications such as text extraction and recognition. Vlahos [12] proposed to key out the foreground objects along with their transparency by exploiting the difference between RGB color channels. Additionally, clustering based methods were also widely adopted in chroma-keying systems. In such chroma-keying systems, pixels in an image are first projected into 3 dimensional color space. According to image color distribution, foreground/background color ranges are defined by some convex closed volumes in the color space. Different techniques were proposed to estimate and represent the shape of foreground/background color ranges [16], [17]. Sampling based method [18] was also proposed to generate transparency matte. In [18], the foreground/background colors are collected by users from user interface. The transparency of a pixel is estimated by looking for the smallest transparency value with all possible foreground/background color combinations.

The aforementioned chroma-keying methods are fundamental in various video editing software or switching systems. However, the existing chroma-keying methods often involve experience based assumptions [12], [16], [17] and manually tuned thresholds [19], which may be not robust in different situations. Moreover, there is little concern on differentiating transparent and reflective regions. In this case, most of the chroma-keying systems require parameters adjustment and postprocessing to obtain satisfactory matting results. As a result, users need to be well trained to generate high quality matte.

In this paper, we propose a coarse to fine procedure to automatically estimate the transparency map of image with little user intervention. Unlike previous methods, which segmented the image without differentiating reflective and transparent regions, we propose a novel segmentation constraint, quadmap, which classifies image pixels with four different types of labels: foreground, background, transparent and reflective regions. By doing this, reflective regions are separated from and treated differently with transparent regions. Therefore, better matting result can be generated because reflective regions can be treated as opaque other than transparent. In order to reliably estimate the background color pixel-wise, we apply Laplacian equation (i.e., harmony equation) to model the background color variation. Considering Laplacian equation in 2D discrete domain, a sparse matrix is constructed to solve the background color. After background color is estimated for each pixel, the transparency factor and the foreground color are then estimated by global sampling and color confidence calculation. The experimental results show that the proposed chroma-keying method can estimate accurate transparency matte and reliable foreground color, especially for images/frames with large transparent or reflective regions such as regions with glass, metal or motion blur.

This paper is organized as follows. In Section II, we present the procedures to label the input image with quadmap.

In Section III, the motivation of background color propagation is introduced along with the sparse matrix solution. The transparency map called α matte is generated based on global sampling and color confidence calculation in Section IV. Section V contains the experiments we have performed. Finally, in Section VI, we draw our conclusion.

II. IMAGE SEGMENTATION AND LABELING BY QUADMAP

In computer graphics, the color $C_{(i,j)}$ of an observed pixel located at (i,j) can be modeled as a linear combination of foreground and background colors using the compositing equation given by

$$C_{(i,j)} = \alpha_{(i,j)} F_{(i,j)} + (1 - \alpha_{(i,j)}) B_{(i,j)} \quad (1)$$

where $F_{(i,j)}$ and $B_{(i,j)}$ are the foreground and background colors respectively, and $\alpha_{(i,j)}$ is the transparency factor which varies from 0 (completely transparent) to 1 (completely opaque).

Foreground objects extraction can also be regarded as pixel-wise α estimation. Given the observed pixel's value $C_{(i,j)}$, the problem of α estimation is highly ill-posed because the number of unknowns (i.e., $\alpha_{(i,j)}$, $F_{(i,j)}$ and $B_{(i,j)}$) is larger than the number of equations. In this case, additional constraints are required to estimate α values as well as the foreground and background colors. One of the most commonly used constraints is called trimap, which can be either manually specified for nature image matting [4]–[6] or automatically generated in chroma-keying systems [17]. Given a trimap, the image pixels are classified into three groups: foreground, background and unknown. The color of pixels in foreground or background group provides clue to solve Equ. (1). The α value of a pixel in unknown group is estimated according to the color of itself, the possible foreground color, and the possible background color. In this paper, the image is further automatically segmented into four different regions as shown in Fig. 1(c): foreground region, background region, transparent region and reflective region. Given an image as shown in Fig. 1(a), the trimap in Fig. 1(b) segments the image into three regions: foreground (white), background (black) and unknown (grey) regions; and our proposed quadmap in Fig. 1(c) segments the image into four regions: foreground (red), background (green), transparent (blue) and reflective (black) regions. Because pixel's color is mixed with background color in both reflective and transparent regions, it is better to differentiate transparency and reflection to avoid false transparency in reflective region. In Fig. 1(d)–(f), the unknown, reflective, and transparent regions are represented by covering other regions with red color. By using our quadmap, the unknown region from conventional trimap is further segmented into reflective and transparent regions. In the following subsections, the separation and labeling of each region will be introduced.

A. Color Distribution of Image to be Chroma-Keyed

Although the tone of background color is already known (i.e., blue or green), the actual background color changes along with studio environment such as light, shadow and

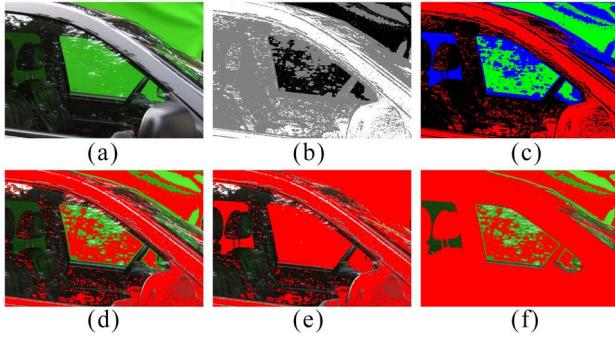


Fig. 1. Comparison of trimap and proposed quadmap. (a) Original image. (b) Trimap. (c) Quadmap. (d) Unknown region obtained from (c). (e) Reflective region obtained from (c). (f) Transparent region obtained from (c).

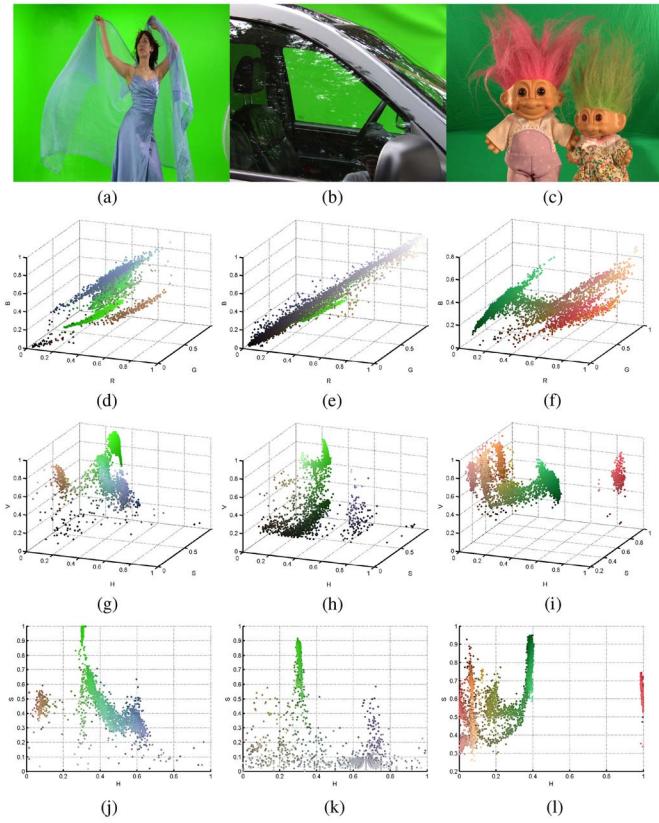


Fig. 2. Color distribution for solid background images. (a) Image "Godiva." (b) Image "Roto." (c) Image "Toy." (d) RGB for "Godiva." (e) RGB for "Roto." (f) RGB for "Toy." (g) HSV for "Godiva." (h) HSV for "Roto." (i) HSV for "Toy." (j) HS plane for "Godiva." (k) HS plane for "Roto." (l) HS plane for "Toy."

rugged background. In this section, images with solid background color are used for analysis. Colors of pixels in such images are plotted in RGB and HSV color spaces as shown in Fig. 2. These two color spaces are representative for two major types of color representations: additive based color space (RGB) and human perception based color space (HSV).

In Fig. 2(d)–(f), it can be observed that similar colors distribute as narrow and long belt from dark to bright in RGB color space. In order to separate foreground from background color in RGB color space, we need to specify a set of convex

closed regions to represent different foreground and background colors. Such regions need to be specified in three dimensional space (i.e., RGB color space) so that the specification involves the estimation of multiple parameters. If HSV color space is used, perceptually similar colors are likely to distribute with similar H values as shown in Fig. 2(g)–(i). This can be more clearly observed if we project colors with different V values to HS plane as shown in Fig. 2(j)–(l). In this case, foreground and background color range can be roughly separated by using only hue values. Compared to RGB color space, it is easier to separate foreground from background color in HSV color space.

B. Background Region Detection Based on Color Statistics and Local Lightness Variation

As mentioned in Section II-A, background region can be detected by thresholding in HSV color space. Fixed thresholds cannot be always appropriate choices because background color distributions are different across images even if their basic tone is the same (e.g., green). In this case, a self-adaptive method based on background color statistics is proposed to estimate the thresholds used for background color determination.

The background color range $R = [T_{min}, T_{max}]$ is adaptively specified by estimating thresholds T_{min} and T_{max} . These two thresholds are estimated mathematically as follows by using image histograms based on the fact that the number of pixels with color in range R is significantly larger than the number of pixels with other colors.

Given one channel of the original image, all pixel values are first normalized to the range $[0, 1]$. M bins with interval r are used to uniformly cover the entire range $[0, 1]$. We denote the bins by their centers' locations as presented in vector $\mathbf{c}^0 = \{c_i^0 | i = 1, 2, \dots, M\}$. The range for each bin is $[c_i^0 - r/2, c_i^0 + r/2]$. The total number of pixels with values falling into each bin is counted as occurrence frequency $p(c_i^0)$. The overall statistical color distribution of original image is represented by vector $P(\mathbf{c}^0) = \{p(c_i^0) | i = 1, 2, \dots, M\}$, which is the histogram of the image. $P(\mathbf{c}^0)$ will dramatically change at bins whose locations are in the range $R = [T_{min}, T_{max}]$. The background color range R is estimated as follows.

First, we define an empty set CR^0 , which is updated in each iteration. The bins to be analyzed is initially \mathbf{c}^0 and updated in each iteration. If we denote the bins to be analyzed in k th iteration as $\mathbf{c}^k = \{c_i^k | i = 1, 2, \dots, N^k\}$, their associated occurrence frequency distribution $P(\mathbf{c}^k)$ is shown in Equ. (2).

$$\begin{aligned} P(\mathbf{c}^k) &= P\left(\mathbf{c}^{k-1} | c_i^{k-1} \notin CR^{k-1}\right) \\ &= \left\{ p(c_i^k) | i = 1, 2, \dots, N^k \right\} \end{aligned} \quad (2)$$

Note that \mathbf{c}^k is always a subset of \mathbf{c}^0 . Given vector $P(\mathbf{c}^k)$, its variance v^k is calculated by using Equ. (3).

$$v^k = \frac{\sum_{c_i^k=r/2 | c_i^k \notin CR^{k-1}}^{1-r/2} (p(c_i^k) - P_{ave}^k)^2}{N^k} \quad (3)$$

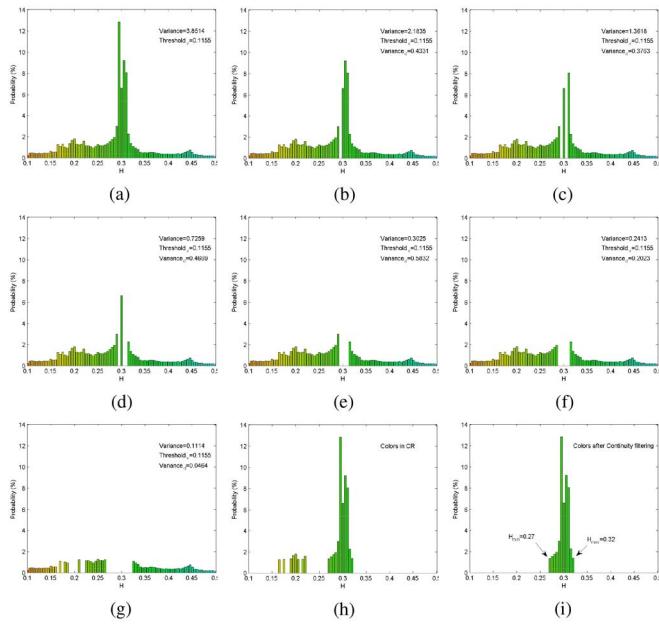


Fig. 3. Histogram analysis for background region detection. (a) Original. (b) 1st iteration. (c) 2nd iteration. (d) 3rd iteration. (e) 4th iteration. (f) 5th iteration. (g) 19th iteration. (h) Candidates. (i) Back color range.

where c_i^k is the values in vector \mathbf{c}^k ; P_{ave}^k is the average of vector $P(\mathbf{c}^k)$; N^k is the number of bins not in set CR^{k-1} ; r is the interval of each bin.

The following two conditions are used to determine when iterations should be ended:

$$v^k < v_{amp} * v^0 \quad (4)$$

$$\frac{|v^k - v^{k-1}|}{v^{k-1}} < T_{grad} \quad (5)$$

where v^0 is the variance of $P(\mathbf{c}^0)$ for original data; v^k is the variance of $P(\mathbf{c}^k)$ in k th iteration; v_{amp} is a constant factor that ensures the current variance is small enough; T_{grad} is a constant value that ensures the variance change between two iterations is small enough.

If the specified conditions are not met, set CR^{k-1} is updated to set CR^k by adding into it the bin c_{max}^k with highest occurrence frequency $p(c_{max}^k)$. Then the estimation proceeds to $(k+1)th$ iteration. If conditions are met at Mth iteration, the final set $CR^{M-1} = \{c_1^{M-1}, c_2^{M-1}, \dots, c_{max}^{M-1}\}$ is used to estimate the thresholds. Given CR^{M-1} , only consecutive bins including the one with highest occurrence frequency are kept to obtain CR^f . Finally, the thresholds are estimated by $T_{min} = \min(CR^f)$, $T_{max} = \max(CR^f)$. In order to demonstrate histogram analysis intuitively, thresholds estimation in H channel of image “Roto” is as shown in Fig. 3. In each figure, “Variance” is the numeric value of v^k in each iteration; “Threshold_v” is the variance threshold calculated from the right side of Equ. (4); “Variance_d” is the numeric value calculated from the left side of Equ. (5) in each iteration. The color of each histogram bin represents the color of the corresponding Hue value. In each iteration, the bin with the highest occurrence frequency is removed until the conditions are met. Fig. 3(h) shows all bins in final set CR^{M-1} . Fig. 3(i) shows consecutive bins CR^f in

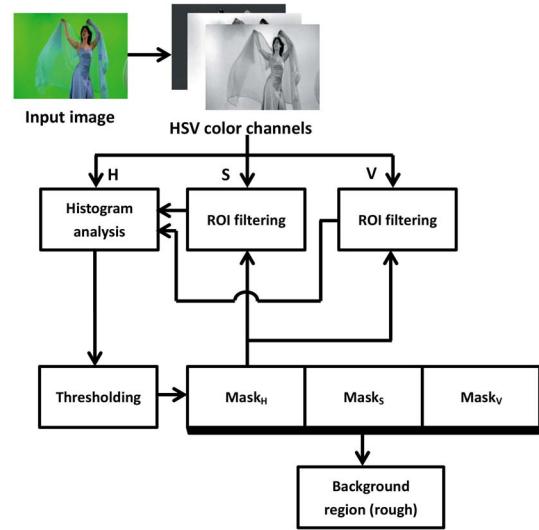


Fig. 4. Flowchart for background detection (rough).

set CR^{M-1} . And the background color thresholds H_{min} and H_{max} in H channel are obtained as shown in Fig. 3(i).

By applying the proposed histogram analysis method to each of the HSV channels, the background region can be extracted. The overall working procedures to roughly segment background region are as shown in Fig. 4. Given an image, the pixel values are transformed to HSV color space. First, the Hue values close to background color (e.g., 0.1-0.5 for green) are histogram analyzed to estimate Hue thresholds H_{min} and H_{max} . A background map $Mask_H$ is consequently generated according to $R_H = [H_{min}, H_{max}]$ by Equ. (6). The S and V channels are respectively ROI (region of interest) filtered according to $Mask_H$. The remaining pixels in S and V channels are also respectively histogram analyzed. The color range (i.e., $R_S = [S_{min}, S_{max}]$ and $R_V = [V_{min}, V_{max}]$) in S and V channels are then obtained to generate $Mask_S$ and $Mask_V$ by Equ. (7) and Equ. (8). And finally, the background mask can be generated by $Mask_{back}(i, j) = Mask_H(i, j) * Mask_S(i, j) * Mask_V(i, j)$.

$$Mask_H(i, j) = \begin{cases} 1 & H(i, j) \in R_H \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$Mask_S(i, j) = \begin{cases} 1 & S(i, j) \in R_S, Mask_H(i, j) == 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$Mask_V(i, j) = \begin{cases} 1 & V(i, j) \in R_V, Mask_H(i, j) == 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

In Fig. 5, the filtering control strength v_{amp} is changed and the background is extracted based on different v_{amp} values. It can be observed that the detected results vary with v_{amp} values. The v_{amp} value used in this paper is set to 0.03 based on tests for large amount of images.

In some situations, background detection solely based on histogram analysis is not enough. It is because some tiny or severely transparent part at the boundary of foreground object may be detected as background as shown in Fig. 6(a). The local lightness variation, which is represented by gradient of V channel, can provide reliable clue to detect such tiny

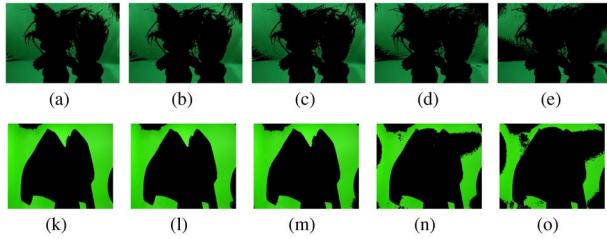


Fig. 5. Background detection with different v_{amp} values. (a) $v_{amp} = 0.01$. (b) $v_{amp} = 0.02$. (c) $v_{amp} = 0.03$. (d) $v_{amp} = 0.10$. (e) $v_{amp} = 0.20$. (k) $v_{amp} = 0.01$. (l) $v_{amp} = 0.02$. (m) $v_{amp} = 0.03$. (n) $v_{amp} = 0.10$. (o) $v_{amp} = 0.20$.

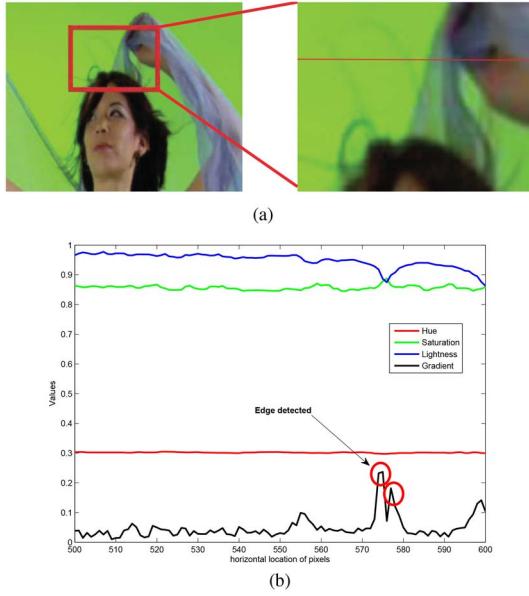


Fig. 6. Detecting tiny part whose color is severely mixed with background. (a) Tiny part at the boundary of foreground object. (b) Color distribution on the horizontal line indicated in image (a).

objects in background as shown in Fig. 6(b). With this concern, the pixels are removed from the background if they generate high gradient in V channel. By doing this, a completely clean background can be obtained as shown in Fig. 7.

C. Foreground and Reflective Region Detection

The foreground region extraction is straightforward because the hue of foreground color is significantly different from the background color. Here we use a predefined factor T_f to control the hue distance between absolute foreground and background. Because Hue values are continuous in a circle ranged from 0 to 1, the foreground color determination involves modulo operation. In order to make the equation clear, a piecewise function is used to determine foreground color range FR , as shown in Equ. (9).

$$FR = \begin{cases} [0, H_b - T_f] \cup [H_b + T_f, 1]; & T_f < H_b < 1 - T_f \\ [H_b + T_f - 1, H_b - T_f]; & H_b > 1 - T_f \\ [H_b + T_f, 1 + H_b - T_f]; & H_b < T_f \end{cases} \quad (9)$$

where H_b is the average Hue in background region.

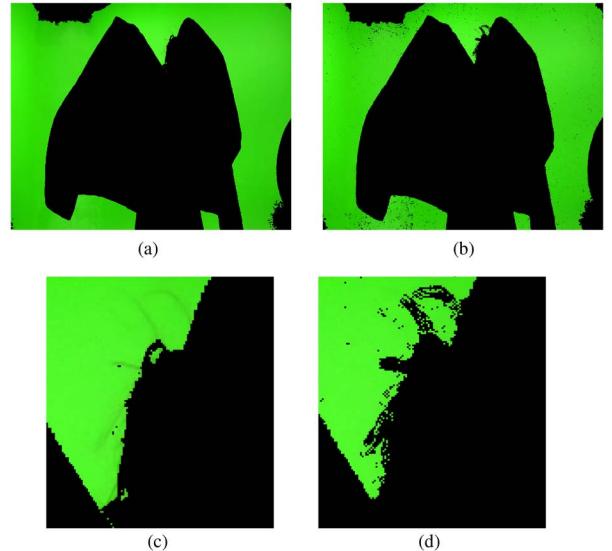


Fig. 7. Clean background by gradient analysis. (a) Extracted background by histogram analysis only. (b) Refined background by gradient analysis. (c) Enlarged part of (a). (d) Enlarged part of (b).

After the foreground and background regions are determined, the remained region is regarded as unknown region in conventional matting schemes. In this paper, the unknown region is further divided into reflective region and transparent region. The reflective region is extracted from unknown region based on two observations from human perception: (1) a region is likely to be reflective if the lightness or saturation is low (Equ. (10)); (2) the hue value of a pixel in reflective region is closer to foreground than to background (Equ. (11)). Observation (1) works for reflective pixel which is over exposed or very dark. Observation (2) works for pixels on object boundary. In this paper, a pixel is regarded as reflective pixel if any of these two conditions in Equ. (10) is met.

$$\begin{cases} S^{ref} < T_{rs} \\ V^{ref} < T_{rv} \end{cases} \quad (10)$$

$$H^{ref} \in [H_{min}^f - \varepsilon, H_{min}^f] \cup [H_{max}^f, H_{max}^f + \varepsilon] \quad (11)$$

where H^{ref} , S^{ref} , V^{ref} are the intensity values of the reflective pixel in HSV color space; T_{rs} and T_{rv} are the saturation and lightness thresholds in assumption (1); ε is the narrow color transition band in assumption (2).

After background, foreground, and reflective regions are detected as described in Section II-B and Section II-C, the remained region is regarded as transparent region in which α value needs to be estimated.

III. BACKGROUND PROPAGATION BASED ON LOCAL AFFINITY

As we known from Equ. (1), it is required to know the background color pixel-wise in order to estimate the foreground color F and transparency value α . In this paper, the background color is estimated by propagating background color from known background region to the whole image based

on local affinity, which is modeled by Laplacian equation as shown in Equ. (12).

$$\Delta f(x, y) = \frac{\partial^2 f(x, y)}{\partial^2 x} + \frac{\partial^2 f(x, y)}{\partial^2 y} = 0 \quad (12)$$

In discrete domain, Equ. (12) refers to that a pixel value equals to the average value of its neighboring pixels. In this case, the background color can be estimated by minimizing Equ. (13).

$$E = \gamma \sum_{i \in \text{back}} (b_i - B_i)^2 + \sum_{i=1}^M \left(b_i - \sum_{j \in N_i} W_{i,j} b_j \right)^2 \quad (13)$$

where i is the index for all pixels in the image; M is the number of all pixels in the image, γ is a large control factor, which ensures that estimated background is close to known background, b_i is the estimated background color; B_i is the known background color in background region; N_i is the neighbors of pixel i ; $W_{i,j}$ is determined according to Equ. (16).

The energy function can be further written in matrix form as

$$E = (B_{\text{est}} - B_{\text{known}})^T \Lambda (B_{\text{est}} - B_{\text{known}}) \dots + ((I - W) B_{\text{est}})^T (I - W) B_{\text{est}} \quad (14)$$

where I is $M \times M$ identity matrix; B_{est} is $M \times 1$ matrix which contains estimated background color; B_{known} has known background color in background region and 0s in other region; Λ and W are both $M \times M$ matrixes and are defined by using Equ. (15) and Equ. (16).

$$\Lambda_{ii} = \begin{cases} \gamma & i \in \text{back} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where γ is a large constant value which guarantees the consistency between calculated background color and known background color.

$$W_{i,j} = \begin{cases} 1/3 & j \in N_i, i \text{ locates at image corner} \\ 1/5 & j \in N_i, i \text{ locates at image side} \\ 1/8 & j \in N_i, i \text{ does not locate at image side} \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

where the denominator is the number of neighbors of a pixel.

Finally, matrix B_{est} can be calculated by using Equ. (17) and the background color is propagated to the whole image as shown in Fig. 8.

$$B_{\text{est}} = [(I - W)^T (I - W) + \Lambda]^{-1} \Lambda B_{\text{known}} \quad (17)$$

IV. α MATTE GENERATION BASED ON GLOBAL SAMPLING

A. Foreground Color Sampling

In Section III, the background color has already been estimated based on propagation. In this case, only foreground color needs to be estimated to calculate α . Conventionally, foreground color is estimated by choosing samples from known foreground regions. Different sampling methods were

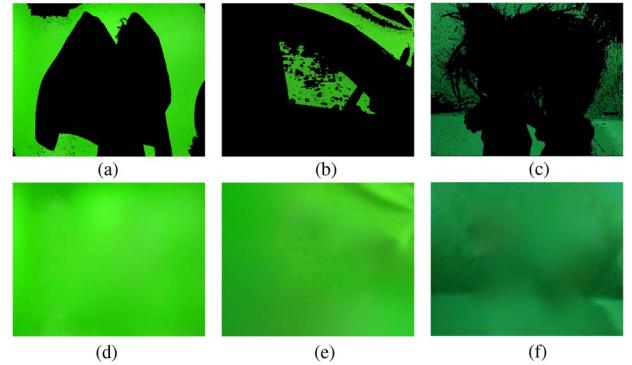


Fig. 8. Background propagation. (a) Back ("Godiva"). (b) Back ("Roto"). (c) Back ("Toy"). (d) Filled back ("Godiva"). (e) Filled back ("Roto"). (f) Filled back ("Toy").

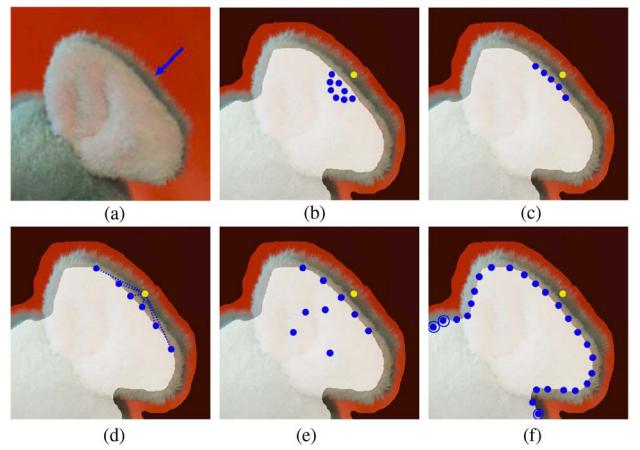


Fig. 9. Different foreground sampling methods. (a) Original image. (b) Nearest sampling. (c) Boundary sampling. (d) Shared sampling. (e) Comprehensive sampling. (f) Global sampling.

proposed as shown in Fig. 9. In Fig. 9 (b)–(f), the bright region refers to foreground region; the dark region refers to background region; the yellow dot represents an unknown pixel whose α value needs to be estimated; the blue dots represent foreground color candidates. In Fig. 9(b), foreground samples are selected from the nearest locations in the foreground region; In Fig. 9(c), foreground samples are selected along boundary between foreground and transparent regions; In Fig. 9(d), a set of rays are drawn from the unknown pixel, and the intersection points of rays and foreground boundary are selected as foreground samples; In Fig. 9(e), foreground samples are not only selected along boundaries but also selected in foreground regions; In Fig. 9(f), foreground samples are globally selected along the foreground boundaries.

In the special case as shown in Fig. 9, a reliable foreground color (i.e., dark grey in this case) cannot be found by using local sampling methods as shown in Fig. 9(b)–(e). Instead, only global sampling with enough searching range can find correct foreground color samples as shown by blue circles in Fig. 9(f).

With this concern, foreground color candidates in this paper are selected in global range as shown in Fig. 9(f). The draw-back for global sampling is the high computational cost. This is always a challenging problem for nature image matting

because the computational complexity is $O(N_f * N_b)$. Here N_f is the number of foreground color candidates; and N_b is the number of background color candidates. The problem, however, is not severe for our chroma-keying method because the background color for each pixel is already known here (refer to Section III). In this case, the computational complexity is reduced to $O(N_f)$, which makes global sampling practical.

B. Determination of Color Confidence

After foreground color candidates are selected, α value can be calculated with each foreground-background color pair by using Equ. (18).

$$\hat{\alpha}_{(x,y)} = \frac{(C_{(x,y)} - B_{(x,y)})(F_{(x,y)} - B_{(x,y)})}{\|F_{(x,y)} - B_{(x,y)}\|^2} \quad (18)$$

After a set of possible α values are calculated for one pixel, it is necessary to determine the confidence about these obtained α values. The α value with the highest confidence will be selected as the final transparency factor of the pixel. The confidence about an α value is determined by the color confidence of the selected foreground and background samples [4]. The color confidence $F(F^i, B^j)$ is defined in Equ. (19).

$$F(F^i, B^j) = \exp\left(-\frac{R_d(F^i, B^j) \times w(F^i)}{\sigma^2}\right) \quad (19)$$

where σ is fixed to 0.1. $R_d(F^i, B^j)$ and $w(F^i)$ are respectively the linear similarity and the color similarity.

The linear similarity is a constraint on the chosen foreground color. A chosen foreground color has higher linear similarity if it better fits the linear relationship with unknown pixel color and background color, as defined in Equ. (20).

$$R_d(F^i, B^j) = \frac{\|C - (\hat{\alpha}F^i + (1 - \hat{\alpha})B^j)\|}{\|F^i - B^j\|} \quad (20)$$

As shown in Fig. 10, the foreground and background pair (F_2, B_2) is a better choice than (F_1, B_1) for the unknown pixel P_c because (F_2, B_2, P_c) have better linear relationship and better fit Equ. (1).

The color similarity is another constraint on the chosen foreground color. A chosen foreground color has higher color similarity if it is closer to the color of the unknown pixel, as defined Equ. (21).

$$w(F^i) = 1 - \exp\left(-\frac{\|F^i - I\|^2}{D_F^2}\right) \quad (21)$$

where D_F is the minimum distance between foreground samples and the unknown pixel in color space.

By choosing the foreground color with the highest color confidence $F(F^i, B^j)$, the optimal α value as well as the optimal foreground color can be estimated for each pixel in the unknown region.

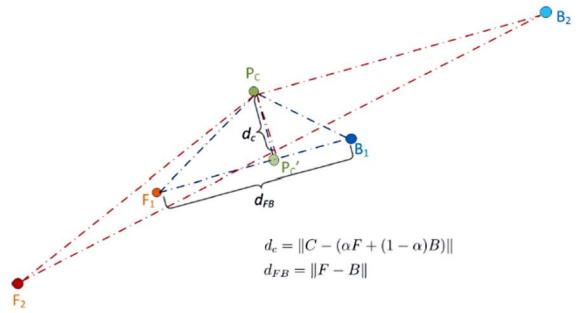


Fig. 10. Linear relationship between foreground, background, and unknown pixel.

V. EXPERIMENTAL RESULTS

In this section, our proposed chroma-keying method is applied to the testing images from two online data resources [20], [21]. Reference [20] provides testing images and evaluation metrics for academic research on α matting. Reference [21] provides support and service for engineers and artists working in the fields of chroma-keying. For all the testing images used in this paper, the color resolutions used in histogram analysis are 0.005 for H channel and 0.01 for S/V channels; T_{grad} in Equ. (5) is set to be 0.3; T_f in Equ. (9) is set to be 0.3. For most of the testing images, T_{rs} and T_{rv} in Equ. (10) are set to be 0.2 and 0.1 respectively.

Given the groundtruth foreground object and groundtruth α map from [20], we generate 16 testing images by composing foreground objects with green background in real scene as shown in Fig. 11(a1)–(a16). Our proposed chroma-keying method estimates the α -maps for the testing images as shown in Fig. 11(b1)–(b16). The foreground objects are extracted and composed on a checkerboard background in order to better present the matting results, which are as shown in Fig. 11(c1)–(c16).

The quality of the generated α -maps are also evaluated by different objective metrics. In order to make comparisons, two other methods are also used to estimate the α -maps. Anat's method [8] was originally proposed to deal with nature image matting. This method performed outstandingly when it was published. It is now still a classical method to deal with nature image matting and many researchers developed new algorithms based on the ideas in this method. "Keylight", which is also used for comparisons here, is one of the state-of-the-art chroma-keying plugins and is integrated in lots of software such as "After Effect", "Nuke" and "Final Cut Pro". This plugin has been used to make lots of films for its good performance. The comparisons are shown in Fig. 12. The testing images are still the 16 images presented in Fig. 11(a1)–(a16). The horizontal axis in Fig. 12 represents the image index numbers, which are the same as they are in Fig. 11. Given the groundtruth α -maps from [20], MSE (mean squared error) and MAE (mean absolute error) of the generated α -maps are calculated. The results are shown in Fig. 12(a) and (b). It can be observed that our proposed method performs well for all the testing images. Meanwhile, it is known [3] that the quality of α -maps can not be always reliably estimated by metrics like MSE or MAE. In this case,

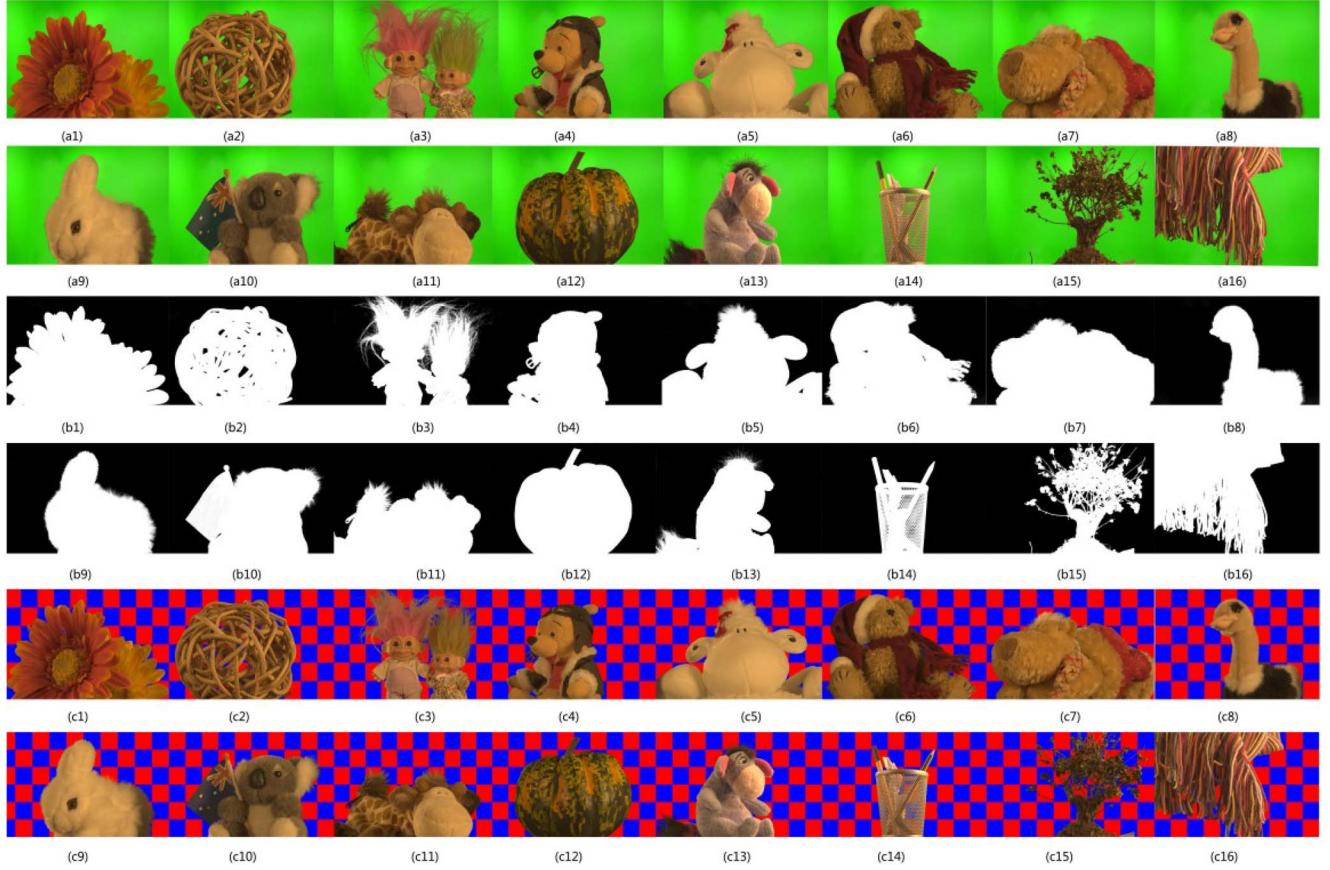


Fig. 11. Chroma-keying for testing images in database [20].

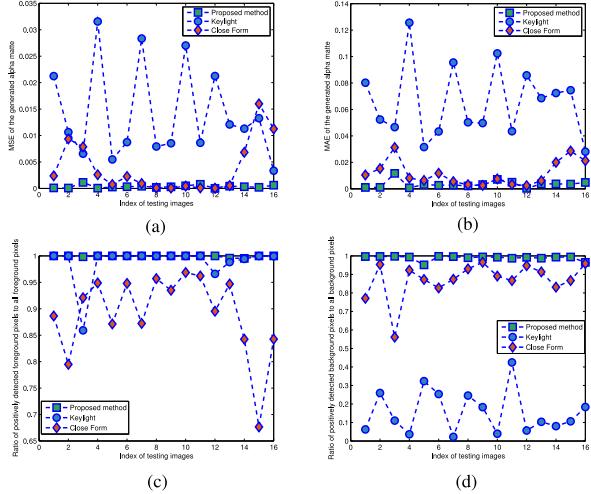


Fig. 12. Alpha matte quality comparisons. (a) Mean squared error. (b) Mean absolute error. (c) Positively detected foreground. (d) Positively detected background.

we use positive detection ratio as supplementary metrics as shown in Fig. 12(c) and (d). Here the positive detection means that an estimated foreground/background pixel is also a foreground/background pixel in the groundtruth alpha map. It can be observed that the positive detection ratio for our proposed keying method is very high for both foreground and background (above 90 %) regions. For foreground region detection, the matting results from “Keylight”

are as good as ours. For background region detection, our method outperforms the other two. This means that our method can reliably extract foreground objects while it can also keep the background clean.

In order to test our method under more challenging conditions, it is applied to five images from [21]. These five images are challenging ones because there are critical situations such as large transparent region, reflective region and motion blurring. By using different matting methods, the estimated α -maps and foreground objects (composed on a checkerboard background) are shown in Figs. 13 and 14. In order to make comparisons, three other methods are used here. Besides Anat’s [8] closed form matting and “Keylight” plugin, we also use Wang’s robust matting [4], a well-known classical local sampling matting method, for comparisons.

Since Wang’s method is based on local sampling, it cannot perform well if reliable foreground color cannot be found in nearby foreground regions. This drawback can be shown in Fig. 13(e1) and (e2). Given a pixel on shirt in Fig. 13(e1), the nearby foreground region is located at the human arm, which cannot provide reliable foreground color candidates for the pixels on the shirt. In this case, the estimated alpha values of most pixels on the shirt are not correct. Anat’s method is based on local affinity. This sometimes leads to unreliable matting result if unknown region is surrounded by only foreground or background regions. As shown in Fig. 13(d1) and (d3), the rear window of the car is incorrectly estimated to be foreground region instead of transparent region because it is completely

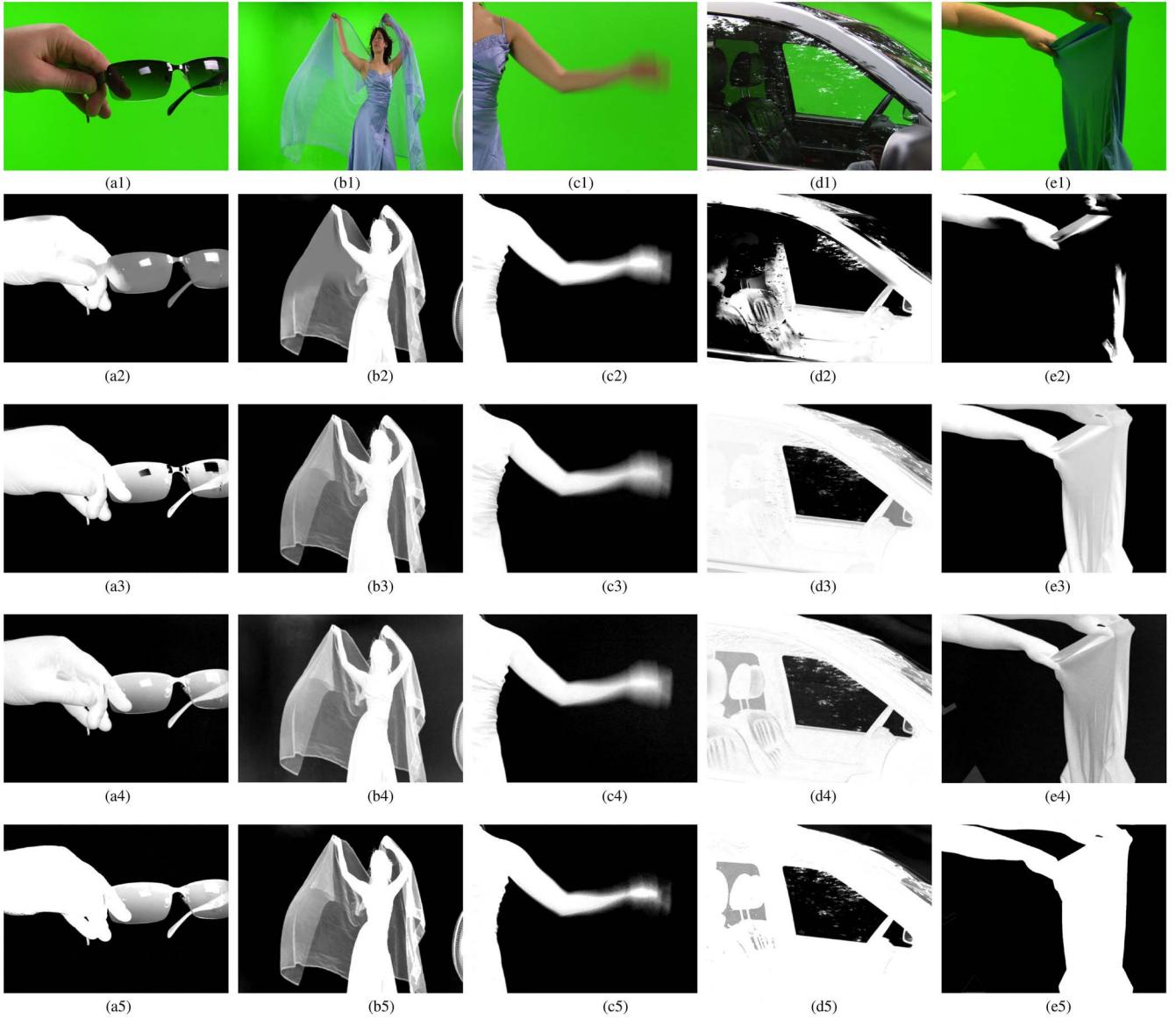


Fig. 13. α maps comparison by using different keying methods. (a1) Image “Glass.” (b1) Image “Godiva.” (c1) Image “Motion.” (d1) Image “Roto.” (e1) Image “Shirt.” (a2) α map of “Glass” generated by Wang [4]. (b2) α map of “Godiva” generated by Wang [4]. (c2) α map of “Motion” generated by Wang [4]. (d2) α map of “Roto” generated by Wang [4]. (e2) α map of “Shirt” generated by Wang [4]. (a3) α map of “Glass” generated by Anat [8]. (b3) α map of “Godiva” generated by Anat [8]. (c3) α map of “Motion” generated by Anat [8]. (d3) α map of “Roto” generated by Anat [8]. (e3) α map of “Shirt” generated by Anat [8]. (a4) α map of “Glass” generated by “Keylight.” (b4) α map of “Godiva” generated by “Keylight.” (c4) α map of “Motion” generated by “Keylight.” (d4) α map of “Roto” generated by “Keylight.” (e4) α map of “Shirt” generated by “Keylight.” (a5) α map of “Glass” generated by our proposed method. (b5) α map of “Godiva” generated by our proposed method. (c5) α map of “Motion” generated by our proposed method. (d5) α map of “Roto” generated by our proposed method. (e5) α map of “Shirt” generated by our proposed method.

surrounded by foreground regions. For the “Keylight” plugin, the background of the generated α -maps is not as clear as what we have in our proposed method. This problem can be clearly observed in Fig. 13(b4) and (e4), which also explains the quality comparisons in Fig. 12(d).

Moreover, our method is outstanding for separating the reflective regions from transparent regions. A reflective region should not be regarded as transparent region. Otherwise, false matting results will be generated as shown in the testing images with reflective regions such as the hand (Fig. 13(a2)–(a4)), the car seats (Fig. 13(d2)–(d4)) and the shirt (Fig. 13(e2)–(e4)). Compared to other methods, our

proposed method estimates the reflective region to be completely opaque (Fig. 13(a5), (b5), (c5), (d5), and (e5)) and restores the original color of the foreground object (Fig. 14(a4), (b4), (c4), (d4), and (e4)) whose color is mixed by reflected background color.

In conclusion, our proposed method performs well for all the testing images, including the standard ones in Fig. 11 and the challenging ones in Fig. 13. It can be shown that our method effectively deals with challenging chroma keying problems such as transparency estimation, reflective region detection, detail (e.g., hair strips) distinguishing and boundary color spill suppression.

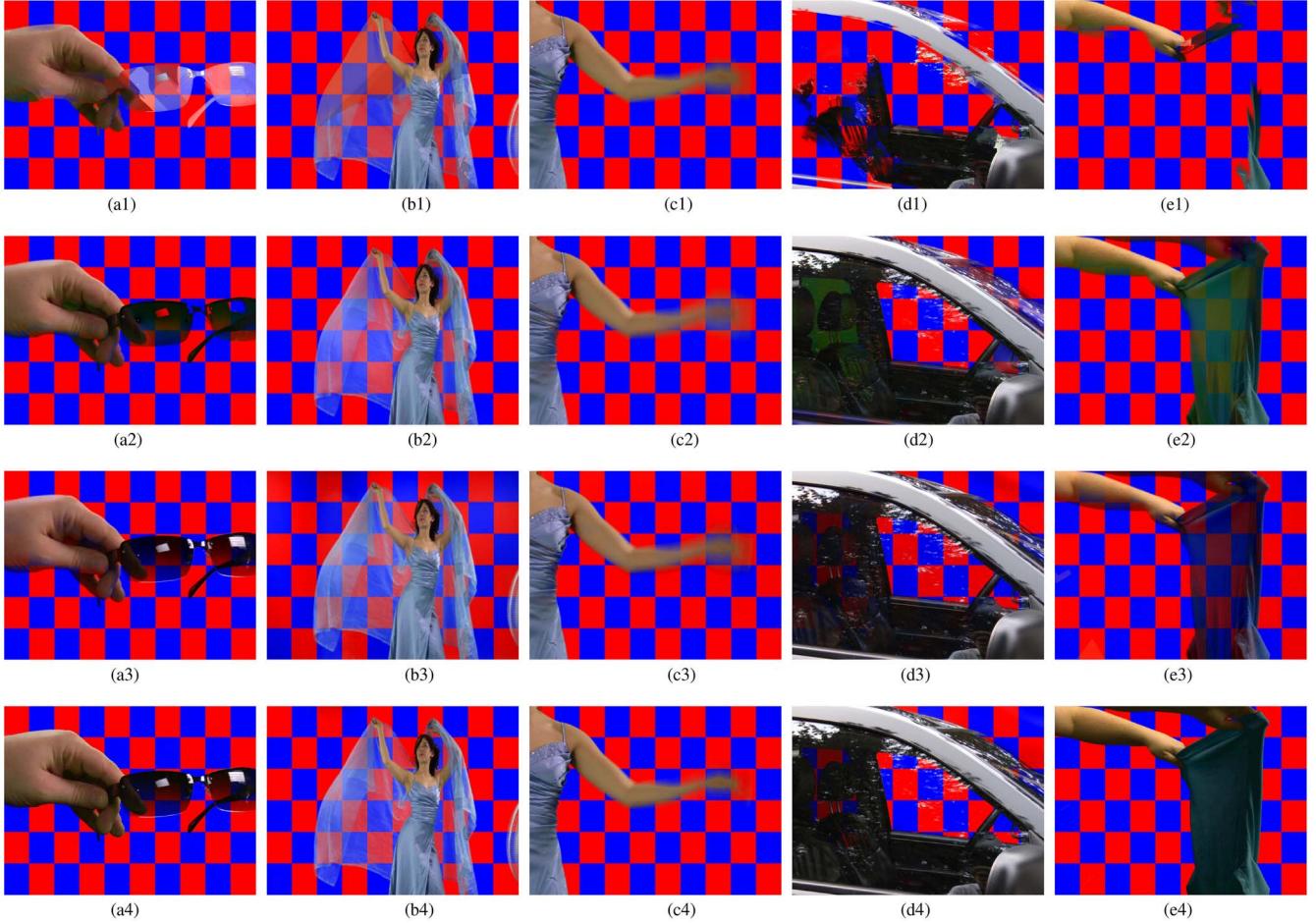


Fig. 14. Composition comparison by using different keying methods. (a1) Composed “Glass” generated by Wang [4]. (b1) Composed “Godiva” generated by Wang [4]. (c1) Composed “Motion” generated by Wang [4]. (d1) Composed “Roto” generated by Wang [4]. (e1) Composed “Shirt” generated by Wang [4]. (a2) Composed “Glass” generated by Anat [8]. (b2) Composed “Godiva” generated by Anat [8]. (c2) Composed “Motion” generated by Anat [8]. (d2) Composed “Roto” generated by Anat [8]. (e2) Composed “Shirt” generated by Anat [8]. (a3) Composed “Glass” generated by “Keylight.” (b3) Composed “Godiva” generated by “Keylight.” (c3) Composed “Motion” generated by “Keylight.” (d3) Composed “Roto” generated by “Keylight.” (e3) Composed “Shirt” generated by “Keylight.” (a4) Composed “Glass” generated by our proposed method. (b4) Composed “Godiva” generated by our proposed method. (c4) Composed “Motion” generated by our proposed method. (d4) Composed “Roto” generated by our proposed method. (e4) Composed “Shirt” generated by our proposed method.

VI. CONCLUSION

In this paper, a novel chroma-keying method is proposed to improve the accuracy and reliability of transparency estimation in chroma-keying system. Given a solid background color image, the color statistics and local lightness variation are respectively analyzed to extract clean background region. Based on human visual perception, the absolute foreground region and the potential reflective region are also segmented. By using these procedures, an image is exclusively segmented into four regions: foreground, background, reflective and transparent regions. Given the known background region and the color in it, background color is smoothly propagated to the whole image by minimizing an energy function, which is derived from the 2D Laplacian equation. The foreground color estimation is based on boundary-wise global sampling with the concern of both requirements for comprehensive color sampling and low computational complexity. Compared to other chroma-keying or α matting methods, reflective region is no longer considered as transparency in our proposed method and

it is processed differently from transparent region. The proposed method can robustly deal with images with background light variation and can significantly improve matting results when there is reflective part on the foreground object.

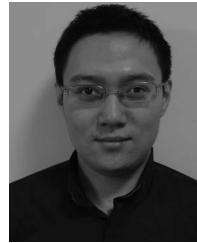
ACKNOWLEDGMENT

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for its generous support.

REFERENCES

- [1] T. Porter and T. Duff, “Compositing digital images,” *ACM SIGGRAPH Comput. Graph.*, vol. 18, no. 3, pp. 253–259, Jul. 1984.
- [2] J. Wang and M. F. Cohen, “Image and video matting: A survey,” *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 2, pp. 97–175, May 2008.
- [3] C. Rhemann *et al.*, “A perceptually motivated online benchmark for image matting,” in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Miami, FL, USA, 2009, pp. 1826–1833.
- [4] J. Wang and M. F. Cohen, “Optimized color sampling for robust matting,” in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Minneapolis, MN, USA, 2007, pp. 1–8.

- [5] K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun, "A global sampling method for alpha matting," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Providence, RI, USA, 2011, pp. 2049–2056.
- [6] E. Shahrian, D. Rajan, B. Price, and S. Cohen, "Improving image matting using comprehensive sampling sets," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, 2013, pp. 636–643.
- [7] Y. Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Kauai, HI, USA, 2001, pp. 264–271.
- [8] A. Levin, D. Lischinski, and Y. Weiss, "A closed form solution to natural image matting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
- [9] X. Chen, D. Zou, S. Z. Zhou, Q. Zhao, and P. Tan, "Image matting with local and nonlocal smooth priors," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Portland, OR, USA, 2013, pp. 1902–1907.
- [10] H. Agata, A. Yamashita, and T. Kaneko, "Chroma key using a checker pattern background," *IEICE Trans. Inf. Syst.*, vol. E90-D, no. 1, pp. 242–249, Jan. 2007.
- [11] B. Vidal, "Chroma key visual feedback based on non-retroreflective polarized reflection in retroreflective screens," *IEEE Trans. Broadcast.*, vol. 58, no. 1, pp. 144–150, Mar. 2012.
- [12] P. Vlahos, "Comprehensive electronic compositing system," U.S. Patent 4 100 569, Jul. 1976.
- [13] S. Shimoda, M. Hayashi, and Y. Kanatsugu, "New chroma-key imaging technique with hi-vision background," *IEEE Trans. Broadcast.*, vol. 35, no. 4, pp. 357–361, Dec. 1989.
- [14] S. Gibbs, C. Arapis, C. Breiteneder, and V. Laliovi, "Virtual studios: An overview," *IEEE Multimedia*, vol. 5, no. 1, pp. 18–35, Jan. 1998.
- [15] R. Brinkmann, *The Art and Science of Digital Compositing: Techniques for Visual Effects, Animation and Motion Graphics*, 2nd ed. Amsterdam, The Netherlands: Morgan Kaufmann, 2008.
- [16] Y. Mishima, "Soft edge chroma-key generation based upon hexoctahedral color space," U.S. Patent U.S.5 355 174 A, Oct. 1994.
- [17] Y. Liu, D. A. Ross, and A. J. Fryer, "Method, system, and device for automatic determination of nominal backing color and a range thereof," U.S. Patent 7 508 455 B2, Mar. 2009.
- [18] A. Berman, P. Vlahos, and A. Dadourian, "Method for removing from an image the background surrounding a selected subject by generating candidate mattes," U.S. Patent 6 288 703 B1, Sep. 2001.
- [19] V. Iverson, "Chroma-key color range determination," U.S. Patent U.S.5 774 191 A, Jun. 1998.
- [20] (2015). *Alpha Matting Benchmark*. [Online]. Available: <http://www.alphamatting.com/>
- [21] (2015). *Data Base for Green Screen Images*. [Online]. Available: <http://www.hollywoodcamerawork.us/greenscreenplates.html>



Wenyi Wang received the B.S. degree in electrical engineering from Wuhan University, China, and the M.A.Sc. degree in electrical and computer engineering from the University of Ottawa, Canada, in 2009 and 2011, respectively. He has been working as a Ph.D. candidate and a Research assistant with the School of Electrical Engineering and Computer Science, University of Ottawa since 2011. His research interests are on image and video processing.



Jiying Zhao received the Ph.D. degrees in electrical engineering from North China Electric Power University and in computer engineering from Keio University. He is a Professor with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada. His research interests include image and video processing and multimedia communications. He is a member of the Institute of Electronics, Information and Communication Engineers and the Professional Engineers Ontario.