MITSUBISHI ELECTRIC RESEARCH LABORATORIES
http://www.merl.com

Rainbow Flash Camera:Depth Edge Extraction Using Complementary Colors

Taguchi, Y.

TR2014-059 May 2014

Abstract

We present a novel color multiplexing method for extracting depth edges in a scene. It has been shown that casting shadows from different light positions provides a simple yet robust cue for extracting depth edges. Instead of flashing a single light source at a time as in conventional methods, our method flashes all light sources simultaneously to reduce the number of captured images. We use a ring light source around a camera and arrange colors on the ring such that the colors form a hue circle. Since complementary colors are arranged at any position and its antipole on the ring, shadow regions where a half of the hue circle is occluded are colorized according to the orientations of depth edges, while non-shadow regions where all the hues are mixed have a neutral color in the captured image. Thus the colored shadows in the single image directly provide depth edges and their orientations in an ideal situation. We present an algorithm that extracts depth edges from a single image by analyzing the colored shadows. We also present a more robust depth edge extraction algorithm using an additional image captured by rotating the hue circle with 180 degrees to compensate for scene textures and ambient lights. We compare our approach with conventional methods for various scenes using a camera prototype consisting of a standard camera and 8 color LEDs. We also demonstrate a bin-picking system using the camera prototype mounted on a robot arm.

International Journal of Computer Vision

Noname manuscript No.
(will be inserted by the editor)

Rainbow Flash Camera: Depth Edge Extraction Using Complementary Colors

Yuichi Taguchi

Received: 28 February 2013 / Accepted: 20 April 2014

Abstract We present a novel color multiplexing method for extracting depth edges in a scene. It has been shown that casting shadows from different light positions provides a simple yet robust cue for extracting depth edges. Instead of flashing a single light source at a time as in conventional methods, our method flashes all light sources simultaneously to reduce the number of captured images. We use a ring light source around a camera and arrange colors on the ring such that the colors form a hue circle. Since complementary colors are arranged at any position and its antipole on the ring, shadow regions where a half of the hue circle is occluded are colorized according to the orientations of depth edges, while non-shadow regions where all the hues are mixed have a neutral color in the captured image. Thus the colored shadows in the single image directly provide depth edges and their orientations in an ideal situation. We present an algorithm that extracts depth edges from a single image by analyzing the colored shadows. We also present a more robust depth edge extraction algorithm using an additional image captured by rotating the hue circle with 180° to compensate for scene textures and ambient lights. We compare our approach with conventional methods for various scenes using a camera prototype consisting of a standard camera and 8 color LEDs. We also demonstrate a bin-picking system using the camera prototype mounted on a robot arm.

Keywords Multi-flash camera · Depth edge extraction · Color multiplexing · Complementary color · Hue circle · Bin picking

1 Introduction

Depth edges (discontinuities) in a scene play a key role in various computer vision applications such as segmentation, 3D reconstruction, object detection, and pose estimation. Using a single 2D image, it is difficult to distinguish depth edges from intensity edges. To robustly detect depth edges, Raskar et al. [16] presented an active illumination camera, called multi-flash camera (MFC), that captures multiple images by casting shadows from different light positions and analyzes the shadow profile in the captured images. In the most basic approach, one light source is turned on at a single capture time.

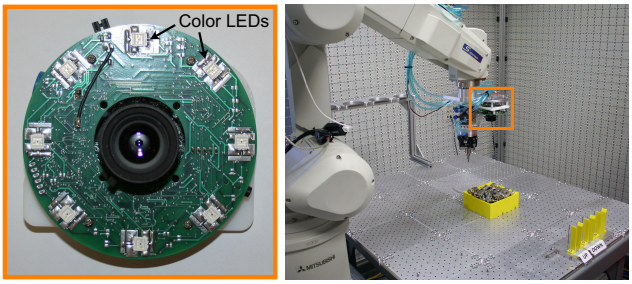

Fig. 1 (a) A prototype of the rainbow flash camera consisting of a standard camera and 8 color LEDs. (b) Our bin-picking system using the rainbow flash camera mounted on a robot arm.

We present *rainbow flash camera*, which multiplexes the conventional MFC imaging to reduce the number of captured images. Instead of flashing each light source at a time, our camera simultaneously flashes all light sources. The multiplexing method is based on the color theory on a hue circle: We exploit the complementary nature of a hue circle to *colorize* shadow regions with different hues depending on the orientations of depth edges, while letting the other non-shadow regions have a neutral (white) color. Our method produces colored shadow regions in a single image, which is sufficient to determine depth edges and their orientations in an ideal situation. We present an algorithm that extracts depth edges from a single image by analyzing the colored

Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA

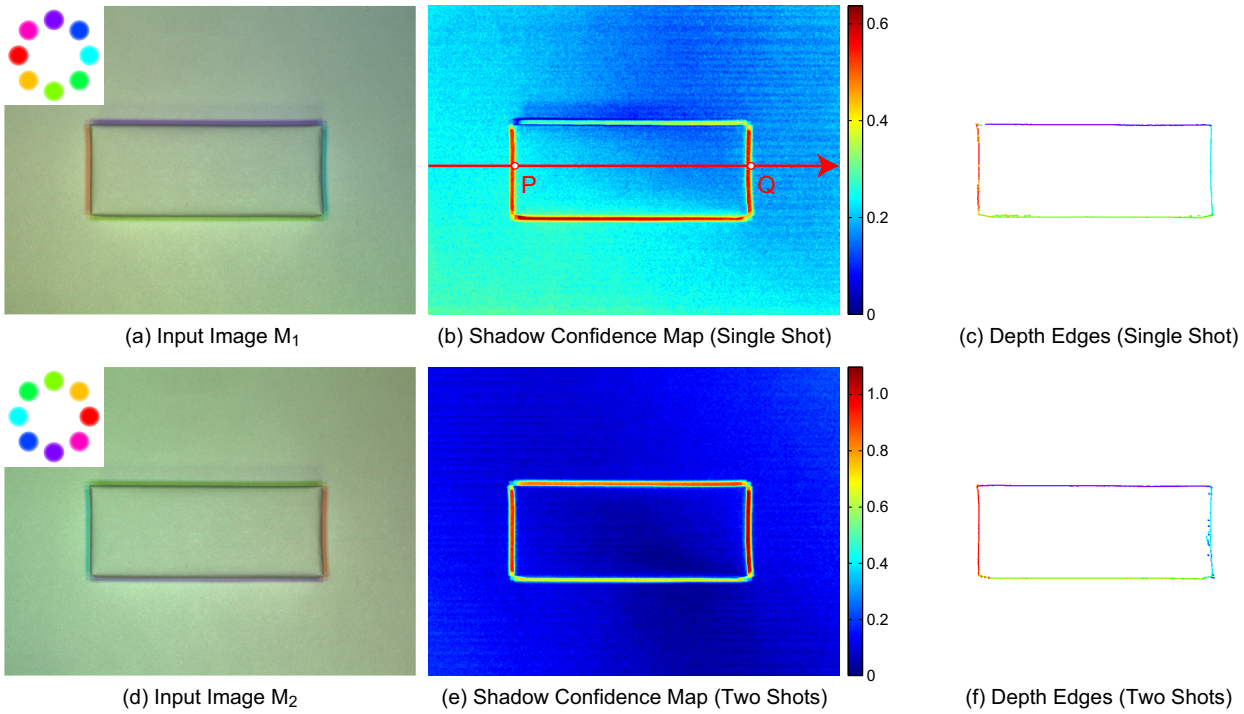

Fig. 2 Overview of our method. (a) Image M_1 captured by simultaneously flashing the 8 color light sources with different hues (shown in the inset). Note that shadow regions are colorized with the different hues, while non-shadow regions have a neutral color. The colors of the shadow regions directly provide the orientations of the corresponding depth edges. (b) Shadow confidence map computed from the single image M_1 (i.e., the saturation component of M_1). An epipolar line along with its traversal direction for the red light source is superimposed on the image. There are two negative transitions P and Q on this line: P is a depth edge for the red light source, while Q is a shadow edge for the cyan light source. These edges can be distinguished by checking the color of the corresponding shadow region. (c) Depth edges extracted from the single image M_1. Each depth edge pixel is depicted with the color of the light source that contributed the most to the depth edge, indicating the edge orientation. (d) For robust depth edge extraction, we capture another image M_2 by rotating the hue circle with $180°$. (e) Shadow confidence map computed from the two captured images, where each pixel value corresponds to the distance between the pixels in the captured images on the hue-saturation plane. (f) Depth edges extracted from the two images.

shadows. We also propose an approach that uses two images, one captured with a hue circle and the other captured by rotating the hue circle with $180°$ (i.e., the complementary version of the original hue circle), to more robustly detect depth edges in the presence of scene textures, non-Lambertian reflections, and ambient lights.

We present a prototype of the rainbow flash camera consisting of a standard camera and 8 color LEDs, as shown in Figure 1(a). We compare the performance of our method with that of conventional methods for various scenes. We also present a bin-picking system, shown in Figure 1(b), that uses the rainbow flash camera mounted on a robot arm to estimate poses of objects randomly placed in a bin and grasp the objects from the bin. Our system replaces the conventional MFC used in [12] with the rainbow flash camera, which greatly reduces the capture and cycle time and enables faster object pickups.

1.1 Contributions

This paper makes the following contributions:

- We present a novel multiplexing method for multi-flash-based depth edge extraction that exploits the complementary nature of a hue circle.
- We describe an algorithm that extracts depth edges from a single image by analyzing the colored shadows, and a more robust algorithm using two images captured with a hue circle and its complementary version.
- We demonstrate a camera prototype using 8 color LEDs and validate our method under various conditions.

A preliminary version of this paper appeared in [20]. Compared to it, we explicitly provide a single-shot depth edge extraction algorithm and discuss its benefits and limitations compared to the two-shot algorithm proposed in [20]. We also demonstrate a bin-picking system using the rainbow flash camera mounted on a robot arm.

1.2 Related Work

Depth Edge Extraction: Raskar et al. [16] introduced a multi-flash camera (MFC) for extracting depth edges by casting shadows from different light positions. Depth edges cor-

respond to discontinuities in the scene geometry and are useful for various applications. Raskar et al. [16] originally used it for non-photorealistic rendering to clearly visualize the scene structure. 3D reconstruction algorithms greatly benefit from depth edges and results using an MFC have been shown for silhouette-based surface reconstruction [4], depth-edge-preserving stereo [6], and depth upsampling [11]. MFC has been also used for robust object detection and pose estimation in heavy clutter in robotic applications [1, 12]. Note that the above works captured multiple images by using one flash at a time.

Feris et al. [8] presented a color multiplexed MFC using three red, green, and blue light sources. They proposed a single-shot method based on learning shadow color transitions for specific scenes, and a two-shot method using an additional reference image captured with white light sources for general scenes. Since their method encodes shadows from the three light sources separately into each RGB channel of the camera, it is only applicable to three light source positions. We present a novel color multiplexing method built on the color theory on a continuous hue circle, thus applicable to any number of light sources that approximate the hue circle. In [22, 3], a single-shot method was presented based on frequency multiplexing. Their method projects multiple sinusoidal patterns such that their frequencies are maintained independent of the scene geometry, and then performs frequency demultiplexing to detect shadow regions [22] or to recover individually illuminated images [3]. Their method requires multiple projectors as light sources, while our method uses multiple color LEDs, resulting in a simpler and more inexpensive system. Their method also sacrifices spatial resolution due to frequency computation in local neighborhoods, while our method enables pixel-wise depth edge extraction as in conventional MFC approaches. Wan et al. [23] proposed single-shot MFC capture as well as other applications using a multi-bucket sensor, which allows multiple images to be captured in a time-interleaved fashion in a single exposure time. Although their approach requires custom image sensor design and fabrication, it could be used complementary to our approach in order to capture two images using different hue circles in a single exposure time.

Illumination multiplexing has been used for various active illumination applications, including photometric stereo [24, 10, 5, 9, 3], structured light [5], image-based relighting [18, 15, 5, 3], object material segmentation [15], as well as depth edge extraction [8, 3]. Color multiplexing was done using standard three (RGB) color channels [24, 8, 10] or more channels with multispectral illumination [15, 9] to reduce the number of captured images, enabling dynamic scene capture or shorter capture process for static scenes. Schechner et al. [18] used multiplexing to improve the signal-to-noise ratio by capturing the same number of images. Chen et al. [3] proposed a frequency multiplexing technique where the frequencies of sinusoidal patterns are maintained independent of the scene geometry as discussed above. De Decker et al. [5] used both color and time multiplexing and showed that at least $\frac{n+2}{3}$ images are required to demultiplex n light sources. The goal of these methods is to demultiplex captured images to obtain multiple images as if the scene were lit by individual light sources. In contrast, our method directly uses the captured images to obtain depth edges *without* demultiplexing. Our method does not have the limitation on the number of light sources; the more discrete light sources we have, the better we can approximate a continuous hue circle.

Complementary colors were used to obtain object material colors in active illumination systems [17, 5]. They captured two consecutive frames by using complementary colors for each projector pixel [17] or each light source [5], and then added the two frames to simulate the case as if the scene were lit by white light sources. Minomo et al. [14] presented a system that projects two images from two projectors such that they have pixel-wise complementary colors on a plane. The system was used for artistic visualization to colorize the shadows of people interacting with the system. We present a novel application of complementary colors for depth edge extraction.

2 Depth Edge Extraction Using Complementary Colors

This section describes the principle of our depth edge extraction method using complementary colors. In this paper, we refer to image regions that are occluded from at least one of discrete light sources or a part of a continuous light source as *shadow* regions, while those lit by all discrete/continuous light sources as *non-shadow* regions.

2.1 Principle in 2D

Let us first explain the principle of depth edge extraction in 2D using two light sources. For now we suppose that there is no ambient light and the scene has a neutral (white) color. We also assume that the scene is Lambertian, and the light sources are located sufficiently far from the scene, illuminating the scene uniformly. As shown in Figure 3(a), a conventional method [16] flashes a single white light source at a time, captures two images, and extracts depth edges as boundaries between shadow (black) and non-shadow (white) regions. There are two such boundaries in this scene, corresponding to a depth edge (depth discontinuity) and a shadow edge (boundary between shadow and non-shadow regions on the same plane). These edges can be distinguished by finding white-to-black transitions in the captured image along the epipolar line defined by the light source position [16].

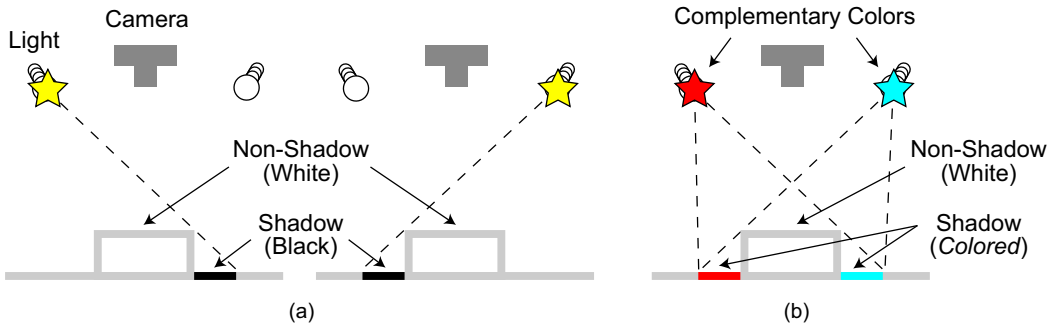

Fig. 3 Depth edge extraction in 2D using two light sources. (a) A conventional MFC flashes a single light source at a time and extracts shadow regions that appear black from each image. (b) Our method simultaneously flashes two light sources having complementary colors. Shadow regions have colors corresponding to the light sources, while non-shadow regions have a neutral (white) color due to the mixture of complementary colors.

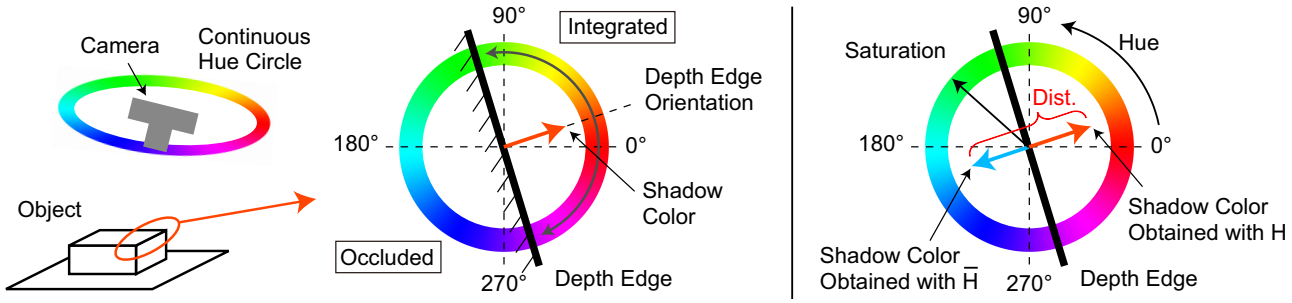

Fig. 4 (Left) Depth edge extraction in 3D using a ring light source having a continuous hue circle. At a depth edge, a half of the hue circle is occluded and the hues in the other half are integrated in the shadow region. This produces colored shadow, whose hue corresponds to the average of the integrated hues. The shadow color directly provides the orientation of the depth edge. On the other hand, all the hues are mixed in non-shadow regions, producing a natural (white) color. Therefore, to extract shadow regions, the saturation component for each pixel (i.e., the distance from the origin on the hue-saturation plane) can be used as a confidence measure. (Right) Robust depth edge extraction using two images. We capture the two images by using a hue circle H and its complementary version $\bar{H}$ and compute the distance between the shadow colors on the hue-saturation plane to obtain a pixel-wise shadow confidence map.

Our method uses complementary colors (red and cyan in Figure 3(b)) for the two light sources and captures a single image by flashing the two light sources at the same time. Because the mixture of complementary colors in an additive color system results in white, non-shadow regions, lit by the two light sources, appear white in the captured image. In contrast, shadow regions, lit by only a single light source due to occlusion, are colorized with the colors of the light sources. Note that the shadow color indicates the orientation of a depth edge.

2.2 Principle in 3D

Here we extend our discussion in 3D. A conventional MFC places N (typically 2, 4, or 8) white light sources around a camera, captures N images by flashing a single light source at a time, and extracts depth edges corresponding to N orientations from each of the captured images. As we use more light source positions, we can distinguish more depth edge orientations in the scene, but we need to take more images.

We build our theory on a continuous light source, instead of discrete light sources. Consider a ring light source around a camera that realizes a continuous, maximally saturated hue circle with the same brightness, as shown in Figure 4 (left). We capture a single image by flashing the entire light source including different hues. In the shadow region corresponding to a straight depth edge of an object[1], a half of the hue circle is occluded by the object, while the other half is integrated in the shadow region.

The additive mixture of any number of colors is determined as the weighted average of the positions of the original colors on the hue-saturation plane, referred to as Newton's geometrical weighting [13]. The brightness of each original color corresponds to the weight. Thus, in our case, the mixture of colors in the half of the hue circle results in a color that has the hue at the center of the half circle and is slightly unsaturated (called saturation cost [13]). As a consequence, shadow regions in the single captured image are colorized according to the orientations of the corresponding depth edges. In non-shadow regions, all colors in the hue circle are mixed, producing a white color. Therefore, shadow regions can be detected as high saturation regions, while non-shadow regions have zero saturation.

[1] For curved depth edges, we consider the tangent line at each depth edge point.

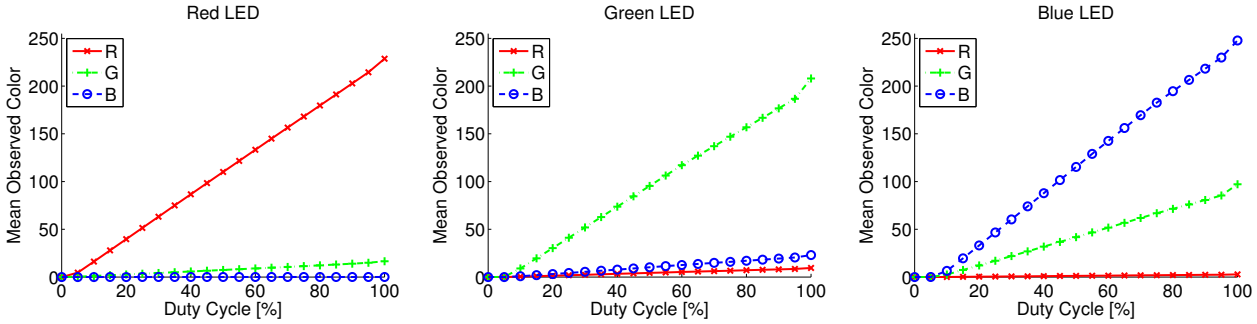

Fig. 5 Colors captured by the camera against the duty cycle for each RGB sub-LED.

Note that in this ideal situation where we use a continuous hue circle and the scene consists only of neutral colors, our method can detect shadow regions and corresponding depth edges, as well as distinguish an infinite number of depth edge orientations, from a single captured image by simply looking up the hues and saturations in the image. In contrast, conventional methods require more images to detect and distinguish more depth edge orientations[2].

In the next section, we present a practical camera implementation by approximating the continuous hue circle using a set of discrete light sources. We show how to extract depth edges from a single image captured with the camera by analyzing the colored shadows. In practice, however, object colors and ambient lights are typically unknown, which confuses the single-shot method. We therefore present a more robust depth edge extraction algorithm that uses two images, one captured using a hue circle and the other captured by rotating the hue circle with $180°$.

3 Implementation

3.1 Camera Prototype

Figure 1(a) shows our camera prototype, consisting of a standard camera (Point Grey Dragonfly) surrounded by 8 color LEDs. Each color LED is composed of three sub-LEDs of RGB components. We control the brightness of the individual RGB components by software pulse-width modulation (PWM) using a PIC microcontroller. All LEDs are synchronized with the camera via a trigger signal sent from the camera to the PIC. We used LEDs that produce illuminations substantially stronger than the ambient illuminations of typical indoor lighting conditions. In experiments, we captured images under such indoor lighting conditions.

Color Calibration: We measured response functions for the RGB components by changing the duty cycle of the PWM signal. We assumed that the 8 RGB components have the same response function for each color channel. We captured multiple images by flashing a single R, G, or B component of all the 8 color LEDs simultaneously with different duty cycles. Figure 5 depicts the mean observed color in the captured image at each duty cycle for each RGB component, showing that our prototype has approximately linear response functions with some color cross talks. To obtain a desired color (r_0, g_0, b_0) for each color LED, we fitted linear functions to the response functions and computed the duty cycles required for each RGB component by solving a linear least-squares problem.

[2] Vaquero et al. [21] showed that three light positions are sufficient to cast shadows for all depth edges in general scenes. However, it only distinguishes three depth edge orientations.

3.2 Depth Edge Extraction Using a Single Image

Here we describe a single-shot depth edge extraction algorithm using the camera prototype. We used a discrete set of hues $H_d = \{45° \times i \mid i = 0, \ldots, 7\}$ with the maximum saturation to approximate a continuous hue circle. The same brightness (value) was set for the 8 HSV colors such that the maximum RGB values stay within the maximum duty cycle obtained in the color calibration process.

Algorithm 1 summarizes our depth edge extraction algorithm using a single image. We define a *shadow confidence map*, which describes the confidence of the shadow for each pixel. For the single-shot algorithm, we simply use the saturation component of the captured image (Step 3), because shadow regions lit by a half of the hue circle should have high saturation colors, while non-shadow regions should have low (ideally zero) saturation, neutral colors, as described in Section 2.2. To implement Step 5, we use a 3×3 Sobel filter whose direction is aligned with the illumination direction. We also use non-maximum suppression for each direction to obtain thin edges that better localize the depth discontinuities.

The shadow color check process in Step 6 is a key to our approach for two different reasons. First, it distinguishes some of intensity edges from depth edges, because depth edges highlighted by our method should have a specific color corresponding to its orientation, while intensity edges can have any color at any orientation. Second, the negative tran-

Algorithm 1 Single-shot depth edge extraction

1: Capture a single image M_1 with the set of hues H_d.
2: Convert M_1 into the HSV color space: $(M_1^{\mathrm{H}}, M_1^{\mathrm{S}}, M_1^{\mathrm{V}})$.
3: Use M_1^{S} (the saturation component of M_1) as a shadow confidence map.
4: **for** each of the N light sources **do**
5: **Epipolar Line Traversal:** Traverse pixels in the shadow confidence map along the epipolar line corresponding to the light position and find step edges with negative transitions.
6: **Shadow Color Check:** For each edge pixel **e**, if the colors of previous k pixels along the traversal direction are close to the color of the target light source, mark the pixel **e** as a depth edge candidate.
7: **end for**
8: For the depth edge candidates, perform connected component analysis and hysteresis thresholding, similar to the Canny edge detector [2].

sitions could be a shadow edge of an antipodal light source. See Figure 2(b)—If we traverse the center row of the image from left to right to find depth edges for the red light source, there are two negative transitions (denoted by P and Q): one of them (P) is a depth edge for the red light source, while the other (Q) is a shadow edge for the cyan light source. In experiments, we checked the following two conditions for $k = 5$ pixels: The saturation should be greater than a threshold of 0.3 and the hue should be within a threshold of $45°$ from the hue of the target light source.

3.3 Robust Depth Edge Extraction Using Two Images

The single-shot approach presented in Section 3.2 will be unstable in the presence of scene textures, ambient lights, and noise. We thus propose a more robust approach that captures two images M_1 and M_2, one by using a hue circle H and the other by rotating the hue circle with $180°$, i.e., using the complementary version $\bar{H}$ of the original hue circle H. As shown in Figure 4 (right), the color of a shadow region obtained with H and the color of the same shadow region obtained with $\bar{H}$ become complementary; i.e., the distance between the two colors on the hue-saturation plane is maximized. We use this distance as the confidence measure to detect shadow regions.

Algorithm 2 describes our two-shot depth edge extraction algorithm. We use a discrete set of hues $H_d = \{45° \times i \mid i = 0, \ldots, 7\}$ and its complementary set $\bar{H}_d = \{180° + 45° \times i \pmod{360°} \mid i = 0, \ldots, 7\}$ to capture the two images. The main difference from the single-shot algorithm is Step 3, where we use the pixel-wise distance between the two images on the hue-saturation plane as the confidence

Algorithm 2 Two-shot depth edge extraction

1: Capture two images M_1 and M_2 with the sets of hues H_d and $\bar{H}_d$, respectively.
2: Convert M_1 and M_2 into the HSV color space: $(M_1^{\mathrm{H}}, M_1^{\mathrm{S}}, M_1^{\mathrm{V}})$ and $(M_2^{\mathrm{H}}, M_2^{\mathrm{S}}, M_2^{\mathrm{V}})$.
3: Compute, for each pixel, the distance between M_1 and M_2 on the hue-saturation plane (Equation 1) to produce a shadow confidence map.
4: **for** each of the N light sources **do**
5: **Epipolar Line Traversal:** Traverse pixels in the shadow confidence map along the epipolar line corresponding to the light position and find step edges with negative transitions.
6: **Shadow Color Check:** For each edge pixel **e**, if the colors of previous k pixels along the traversal direction are close to the color of the target light source in the two images, mark the pixel **e** as a depth edge candidate.
7: **end for**
8: For the depth edge candidates, perform connected component analysis and hysteresis thresholding, similar to the Canny edge detector [2].

measure, instead of the saturation component of the single image. Specifically, for each pixel $\mathbf{x}$, we compute

$$D_{\mathrm{HS}}(M_1, M_2)(\mathbf{x}) = \left\| M_1^{\mathrm{S}}(\mathbf{x}) \begin{pmatrix} \cos(M_1^{\mathrm{H}}(\mathbf{x})) \\ \sin(M_1^{\mathrm{H}}(\mathbf{x})) \end{pmatrix} - M_2^{\mathrm{S}}(\mathbf{x}) \begin{pmatrix} \cos(M_2^{\mathrm{H}}(\mathbf{x})) \\ \sin(M_2^{\mathrm{H}}(\mathbf{x})) \end{pmatrix} \right\|, \quad (1)$$

where M_i^{H} and M_i^{S} $(i = 1, 2)$ are the hue and saturation components of M_i ($0° \le M_i^{\mathrm{H}}(\mathbf{x}) < 360°$, $0 \le M_i^{\mathrm{S}}(\mathbf{x}) \le 1$). Note that intensity edges due to scene textures and ambient lights are removed from the shadow confidence map, because the distance measure becomes zero for these intensity edges. Step 6 is still required to distinguish shadow edges from depth edges, but this can be done more easily than the single-shot approach, where intensity edges as well as shadow edges should be distinguished from depth edges. In experiments, therefore, we only checked the closeness of the hue (a threshold of $90°$) for $k = 1$ pixels.

Note that instead of capturing the second image with $\bar{H}$, one could capture an ambient image (no light source on) or an image with making the colors of all light sources white [8]. However, they produce neutral color illuminations, corresponding to the center of the hue-saturation plane. Thus the distance measure becomes less reliable than our method. We will validate this in Section 4.2. Any other rotation of the original hue circle H also produces a less reliable distance measure, because it makes the angle between the two shadow colors less than $180°$, while the saturations of the shadow colors are always the same due to the integration of a half circle.

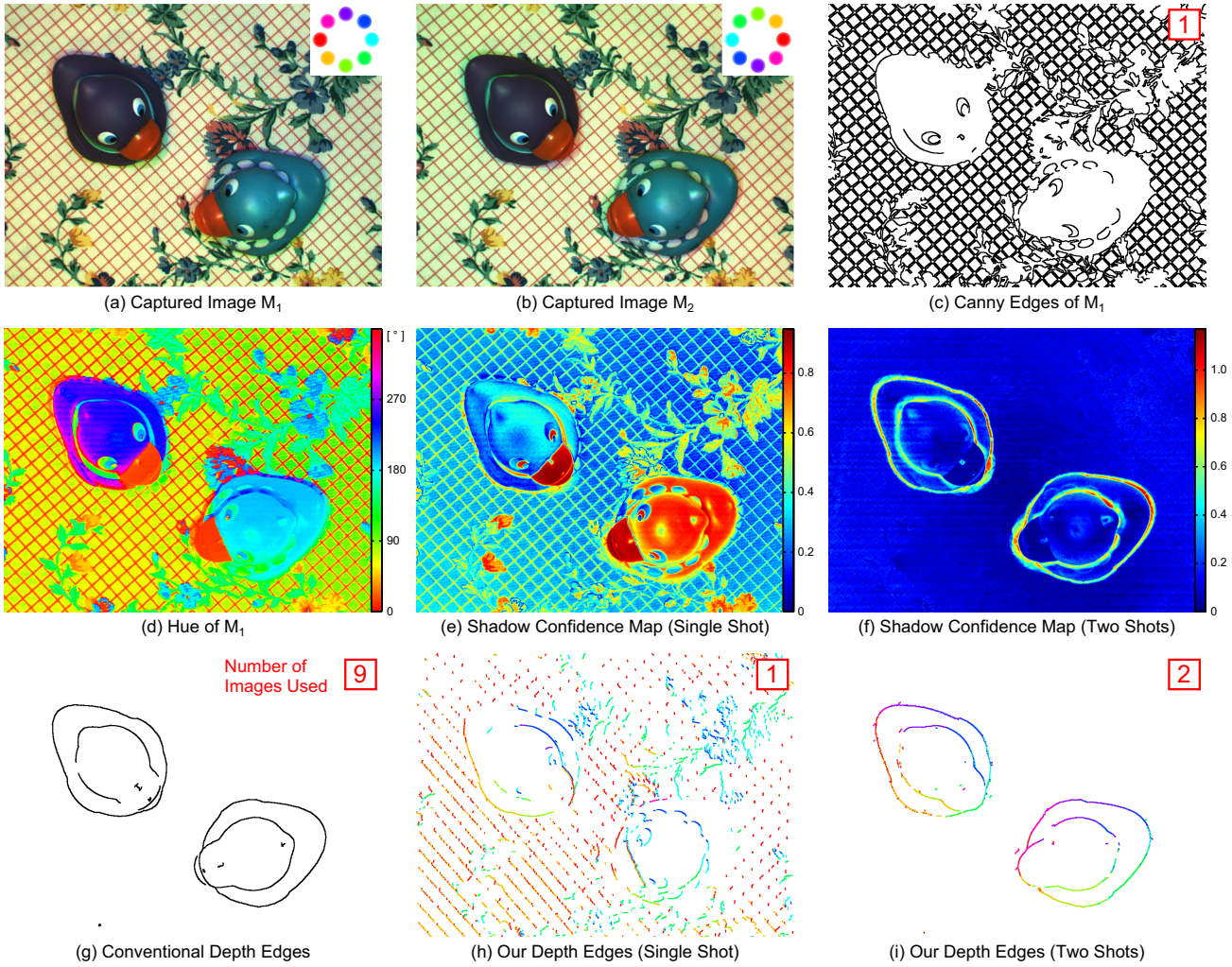

Fig. 6 Results for a highly textured scene. (a, b) Captured images M_1 and M_2. (c) Edges obtained by applying the Canny edge detector [2] to M_1. Note that we used relatively high thresholds to avoid too many intensity edges; the result still includes many intensity edges, while some of depth edges are missing. (d) The hue component of M_1. (e) Shadow confidence map using a single image M_1 (i.e., the saturation component of M_1). (f) Shadow confidence map using the two images. (g) Conventional depth edges obtained using 9 images. (h) Our depth edges obtained using a single image M_1. Note that the result includes less intensity edges than that using the Canny edge detector shown in (c) because of the shadow color check step. (i) Our depth edges obtained using the two images, which are comparable to those obtained using the conventional method.

4 Experiments

In this section, we evaluate our method under various conditions by comparing it with the conventional MFC method [16] and Feris et al.'s RGB multiplexing method [8]. We also demonstrate a bin-picking system using the rainbow flash camera and compare it with the one using the conventional MFC [12].

For visualization purpose, we color-coded our depth edge results with the color of the light source that provided the maximum Sobel filter response. Thus the color indicates the depth edge orientation. Note that the conventional method also extracts depth edges with N orientations; however, we depicted them with black color without showing the orientations in the results shown in this paper. We also dilated the extracted depth edges by one pixel for better visibility.

4.1 Comparisons with the Conventional Method

To obtain conventional MFC images, we turned all RGB components of a color LED on to produce a white light source and captured N flash images, $I_1, \ldots, I_N$. The conventional method also requires one ambient image, A, to compensate for object colors and ambient lights. To obtain depth edges, we performed the following steps [16]:

1. Subtract the ambient image from flash images: $I_i^s = I_i - A$ $(i = 1, \ldots, N)$.
2. Compute a pixel-wise maximum image to generate a *shadow-free* image: $I_{\max}(\mathbf{x}) = \max_i(I_i^s(\mathbf{x}))$.

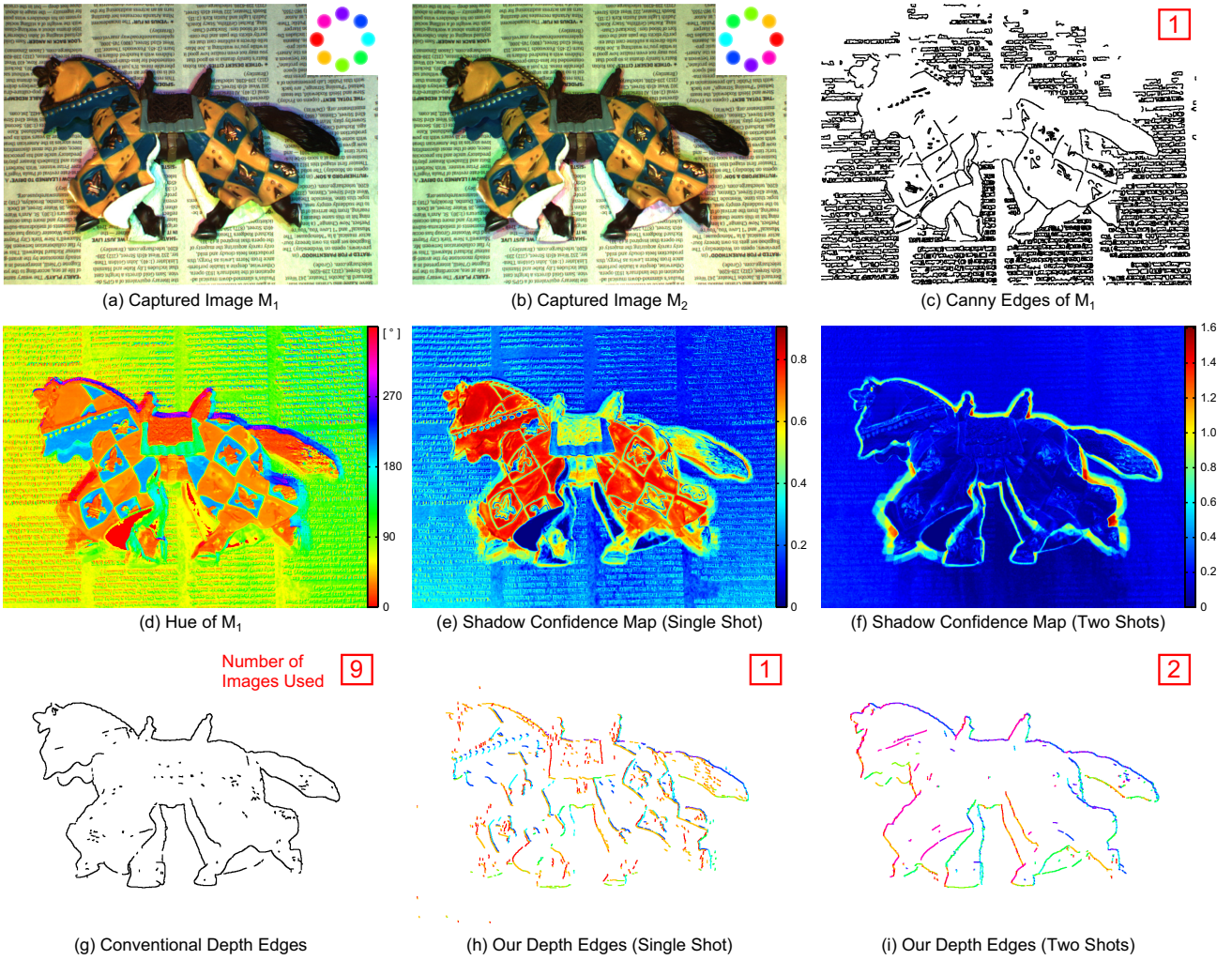

Fig. 7 Results for another highly textured scene. See the caption of Figure 6 for the details.

3. Create a ratio image for each flash position: $R_i(\mathbf{x}) = I_i^s(\mathbf{x})/I_{\max}(\mathbf{x})$.

The ratio images R_i act as shadow confidence maps for each light position in the conventional method. We generated conventional depth edges by applying the same edge detection algorithm used in our method (Steps 4 to 8), except the shadow color check (Step 6), to the ratio images.

Scene Textures: Figures 6 and 7 compare results obtained using different methods for highly textured scenes. The standard Canny edge detector [2] extracts both depth edges and intensity edges for such scenes. Our single-shot method also extracts some intensity edges, since the shadow confidence map computed from a single image becomes incorrect due to high saturation colors of object textures. However, note that our single-shot method only extracts edges where the hue in a high saturation region matches to that of a light source corresponding to the edge orientation; for example in Figure 6, our single-shot method extracts only one of the four diagonal edge orientations of the check pattern in the background, while the Canny edge detector extracts all the four edge orientations. Our method using two images produces robust shadow confidence maps even in the presence of scene textures, and produces comparable depth edges to the conventional method that requires 9 images. In the following results, we focus on the two-shot approach due to its robustness.

Note that our method requires shadow regions to reflect all the colors of the light sources; otherwise it will miss depth edges corresponding to the orientations of colors that are not reflected. In contrast, the conventional method requires shadow regions to reflect some colors (spectrums) of white light sources; thus it is more robust to object colors.

Number of Flashes: Figure 8 shows the effect of using different numbers of flashes for depth edge extraction. For our method, we used a reduced set of color light sources to capture the input images, as shown in the insets of Figure 8. For the conventional method, we used the same reduced set of the light source positions as our method. To extract depth

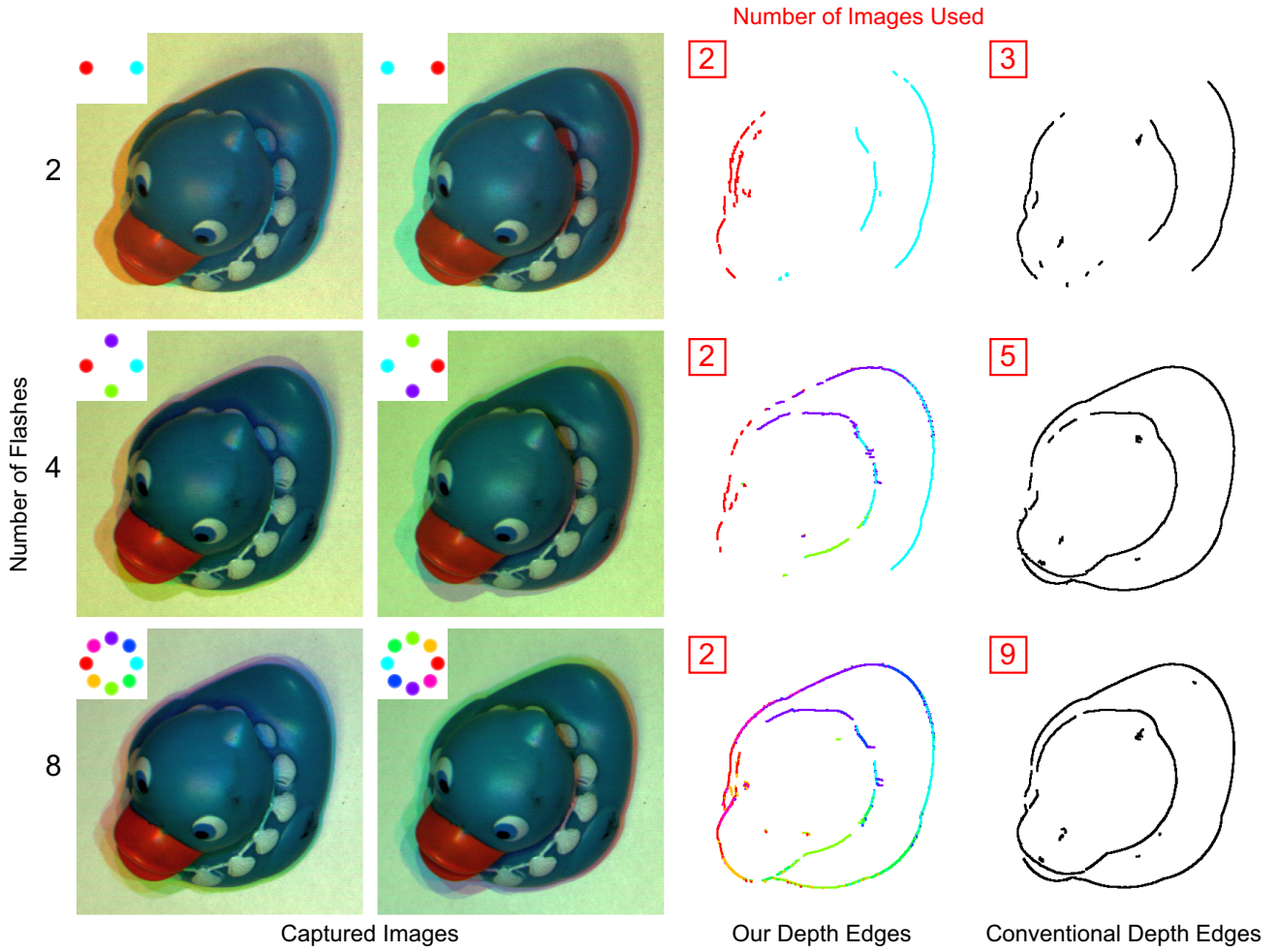

Fig. 8 Depth edge extraction using different numbers (2, 4, and 8) of flashes. For our method, we used a reduced set of color light sources as shown in the insets to capture the two images. Note that the shadow regions in the captured images include more colors as we increase the number of color light sources. For the conventional method, we used the same reduced set of the light source positions as our method. As we increase the number of flashes, both methods extract more depth edges at different orientations. Note that the conventional method requires $N + 1$ images, while our method always uses 2 images independent of the number of flashes.

edges using a reduced set of flashes, we made the number of directions used for the epipolar line traversal (Step 5) the same as the number of flashes used. Both methods extract a larger number of depth edges when we use more flashes. Note that the conventional method requires $N + 1$ images to use N light sources, while our method always uses a fixed number (two) of images independent of the number of light sources. Using a larger number of color light sources, our method can better approximate a continuous hue circle, which helps obtain more depth edge orientations using the same number of captured images.

Robustness: There are several effects that violate the assumptions in our method as well as in the conventional method, such as non-Lambertian reflections and global illumination effects (e.g., inter-reflections). Figure 9 shows results for objects having semi-specular surfaces, demonstrating that our method can tolerate a small amount of such violations similar to the conventional method. Note that the hues in non-shadow regions (object surfaces) change in the two captured images mainly due to specular reflections. Such object surfaces also increase the dynamic range of the scene; depending on the normals of the surfaces, some regions reflect a large amount of light and become bright, while other regions remain dark. Both of these too bright or dark regions lead to low saturation colors that are close to the center of the hue-saturation plane, where the estimation of the hue becomes unstable. As shown in Figure 9(e), the absolute difference of the hues would be unstable to be used as the shadow confidence map for such regions. Our shadow confidence map (Figure 9(f)) defined by the distance on the hue-saturation plane provides a more robust measure, because the distance between low saturation colors becomes small. Our depth edges (Figure 9(h)) are comparable to the conventional depth edges (Figure 9(g)). We applied a pose estimation algorithm using the fast directional chamfer matching (FDCM) [12] to our depth edges. Figure 9(i) demonstrates

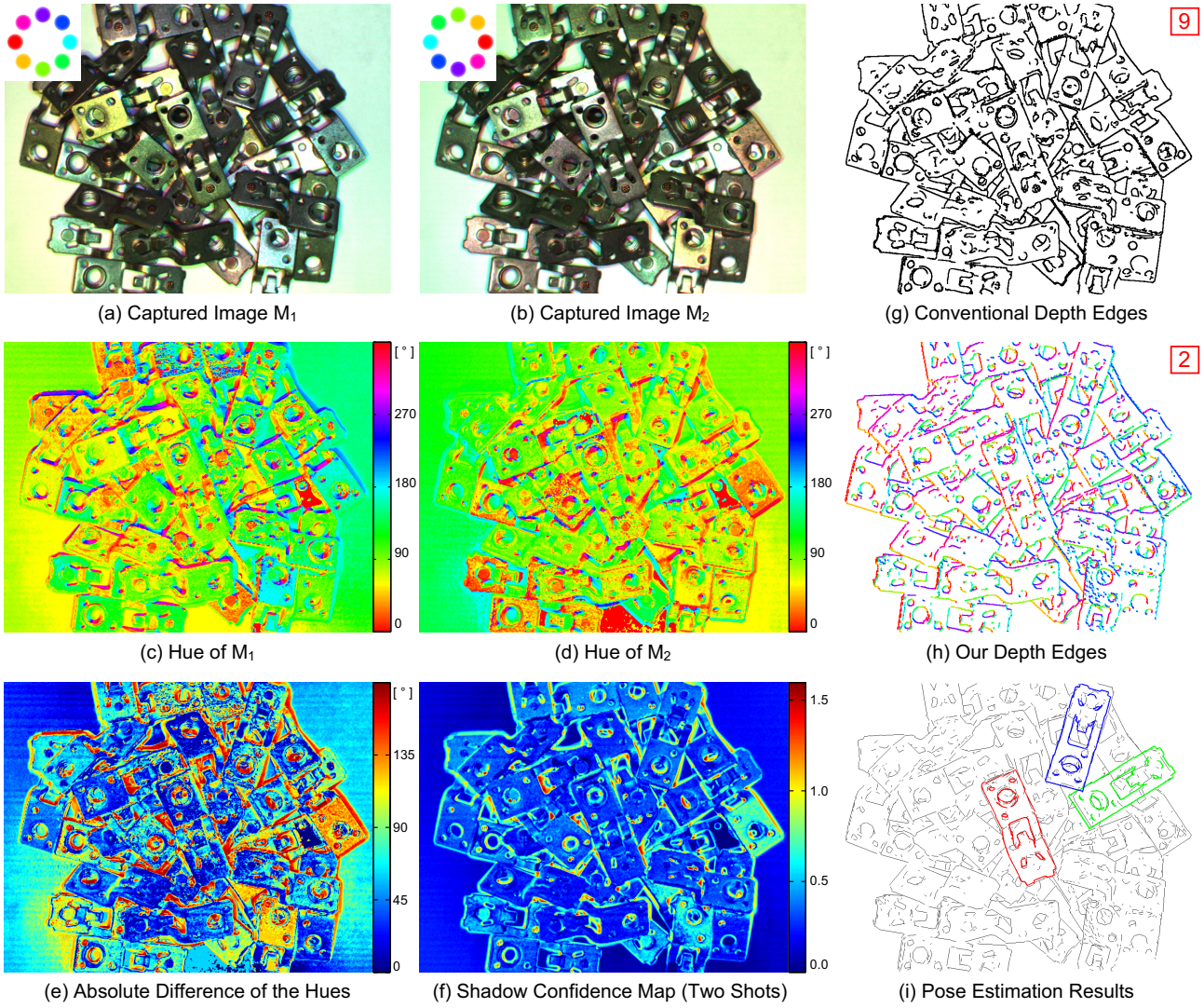

Fig. 9 Results for semi-specular objects. (a, b) Captured images. (c, d) The hue components of the captured images. Note that non-Lambertian reflections on object surfaces violate the assumption that non-shadow regions have a fixed color in the two images. (e) Only using the absolute difference of the hues does not provide a robust shadow confidence map, because colors on semi-specular object surfaces have small saturations, at which the estimation of the hue becomes unstable. (f) The distance on the hue-saturation plane provides a better shadow confidence map. (g) Depth edges extracted with the conventional method, requiring 9 images. (h) Depth edges extracted with our method using 2 images. (i) Object detection and pose estimation results using the fast directional chamfer matching (FDCM) algorithm [12] for our depth edges. Best three estimated poses (red, green, and blue) are superimposed on the line representation of the depth edges.

that we can correctly estimate the poses of objects using our depth edges.

Strong specular highlights (Figure 10) violate our assumption that all the hues are mixed in non-shadow regions, resulting in a neutral color. They also violate the assumption in the conventional method that non-shadow regions have the same appearance in all the flash images. The conventional method, however, can exploit the fact that specular highlights change their positions in the flash images, since each image is captured with a single light source at a different position. In [16, 7], instead of using a pixel-wise maximum value to compute the shadow-free image, the median intensity or median of gradients at each pixel was used to remove the effect of specular highlights. Here we simply used the median intensity to compute the shadow-free image and obtained the depth edges shown in Figure 10 for the conventional method, which do not include artifacts from specular highlights. On the other hand, our method uses all the light sources simultaneously. This produces specular highlights at fixed positions in the two captured images, causing false depth edges.

It has been shown that specular highlights produced by light sources at different positions remain in a small neighborhood for high curvature regions having normals toward the camera [19]. It can be observed in Figure 10(c), where specular highlights corresponding to all the 8 light sources

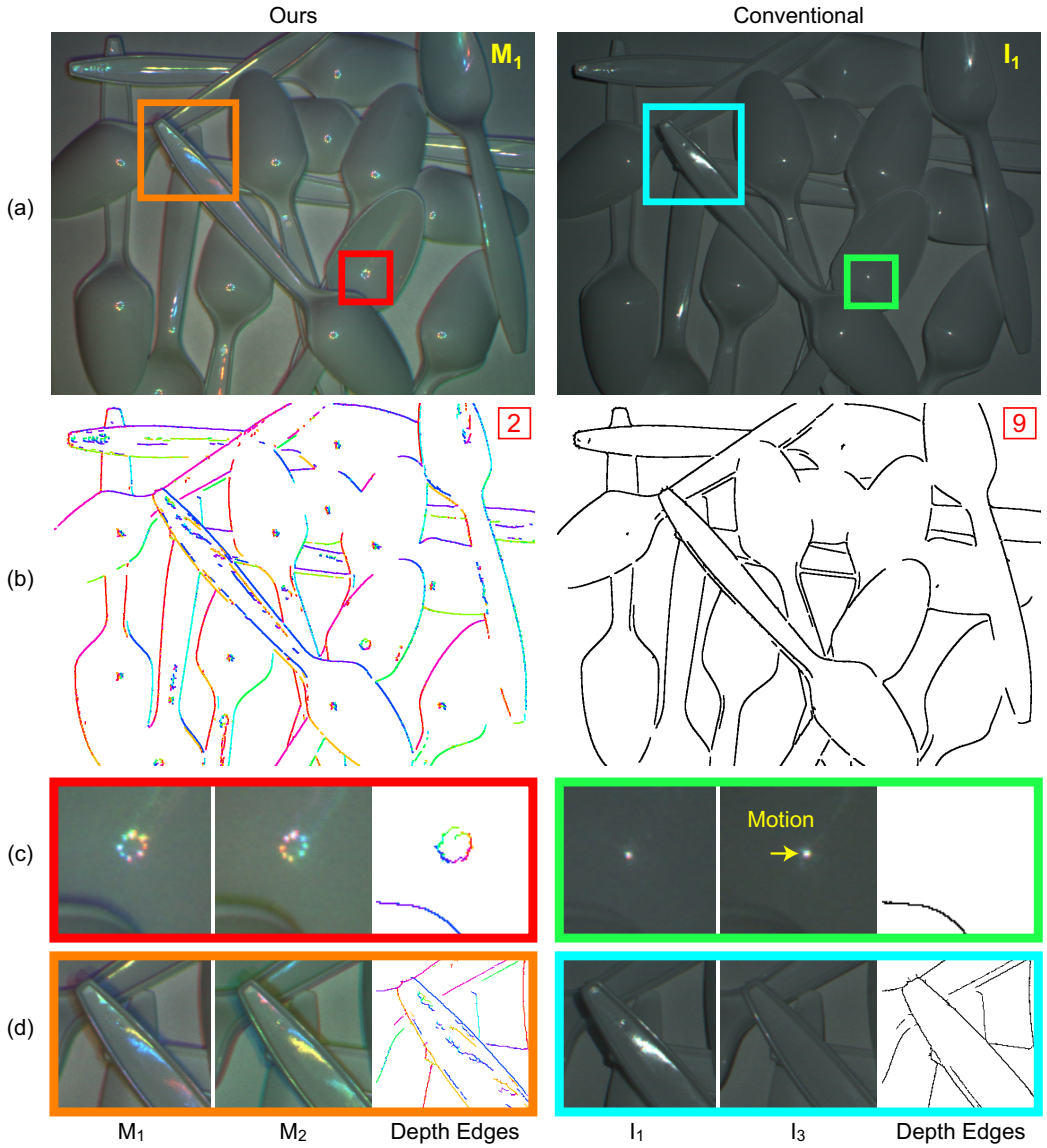

Fig. 10 Effects of specular highlights for (left) our method and (right) the conventional method. (a) One of the captured images. (b) Extracted depth edges. (c, d) Close ups of two captured images (M_1 and M_2 for our method, I_1 and I_3 for the conventional method) and depth edges. In conventional MFC images captured by using a single light source at a time, the positions of specular highlights typically change depending on the positions of the light sources. The conventional method [16, 7] exploits this fact to remove the effect of specular highlights. On the other hand, our method uses all light sources simultaneously, thus producing specular highlights at fixed positions in the two captured images. This confuses our algorithm and generates false depth edges.

appear in a small neighborhood. This special case (all the light source colors appearing in a small neighborhood) could be detected and removed from the depth edge extraction process. However, other cases such as the one in Figure 10(d) would be difficult to be detected.

Processing time of our method was almost the same as that of the conventional method, since both methods use similar building blocks for depth edge extraction. To extract depth edges using 8 light sources from 1024×768 pixel images, both methods took about 200 milliseconds using a C++ implementation on a standard PC with an Intel Core i7-950 processor.

4.2 Comparisons of Different Strategies for Capturing Multiple Images

Our method captures two images M_1 and M_2, M_1 with a hue circle H and M_2 with its complementary version $\bar{H}$. Here we clarify the advantages of our method over other strategies for capturing the second image: (1) capturing an ambient image A, and (2) capturing an image W by making all the light sources white [8], as we discussed in Section 3.3.

Figure 11 compares shadow confidence maps for three different scenes computed using Equation 1 with different pairs of images: (a) M_1 and A, (b) M_1 and W, and (c) M_1

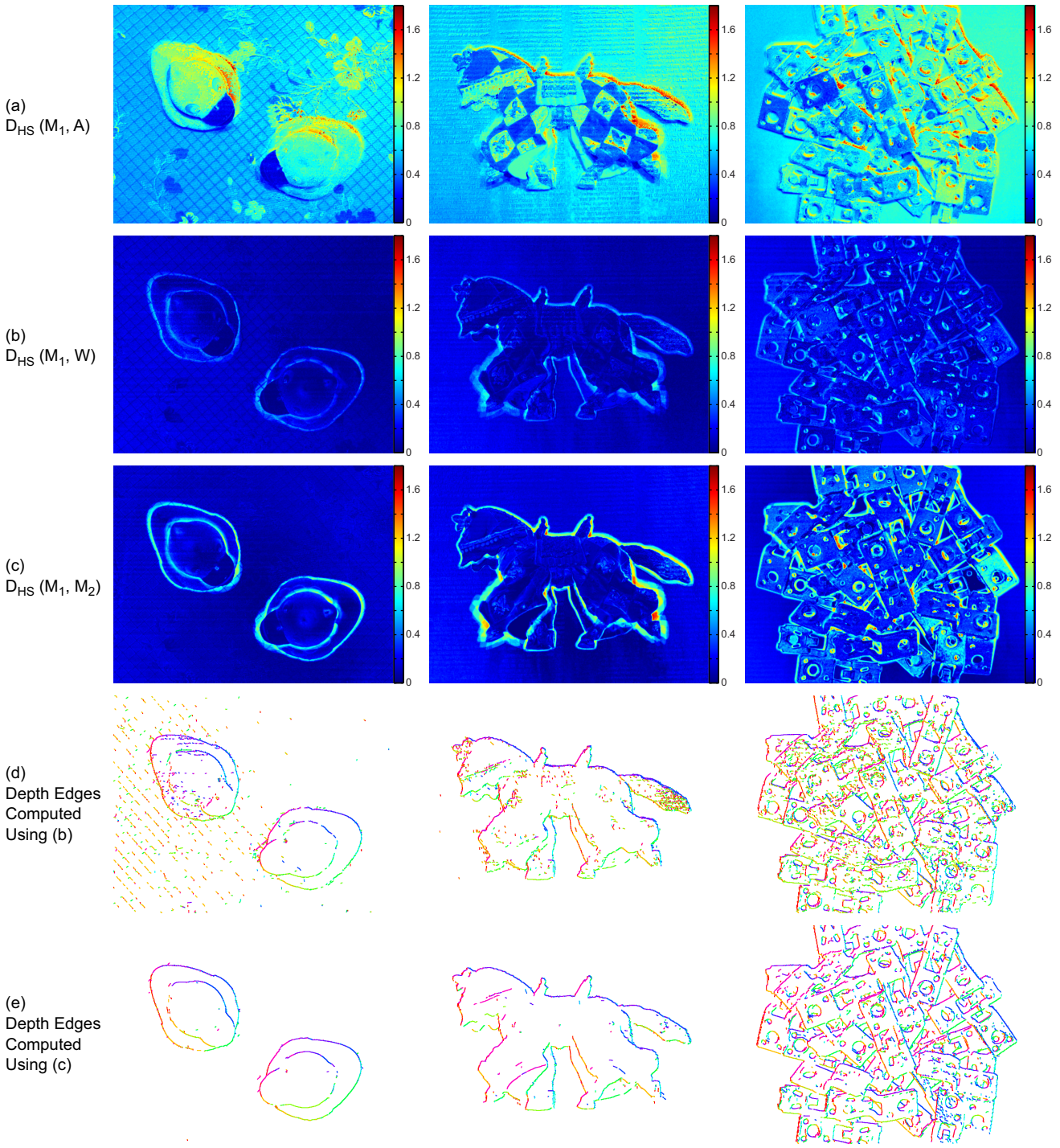

Fig. 11 Comparisons of different strategies for capturing the second image. (a–c) Shadow confidence maps computed using Equation 1 with (a) M_1 and A (ambient image), (b) M_1 and W (image captured by making all the light sources white), and (c) M_1 and M_2 (our method) for three different scenes shown in Figures 6, 7, and 9. Our method using M_1 and M_2 produces the best shadow confidence maps with the highest contrast as shown in (c), because the colors of shadow regions become complementary in the two images. A is generally too dark to compute the hue and saturation, resulting in the poor performance shown in (a). W is captured with neutral color illuminations, which correspond to the center of the hue-saturation plane and lead to lower contrast results shown in (b). (d) Depth edges computed using $D_{\mathrm{HS}}(M_1, W)$ include more false edges than (e) depth edges computed using $D_{\mathrm{HS}}(M_1, M_2)$. This is because (1) $D_{\mathrm{HS}}(M_1, W)$ has lower contrast than $D_{\mathrm{HS}}(M_1, M_2)$, and (2) the shadow color check can be done only using the single image M_1 (W does not include any information on the shadow colors). Note that the shadow confidence maps in (a–c) are depicted using the same scale; thus they are visualized with slightly different colors compared to those shown in Figures 6(f), 7(f), and 9(f). Our depth edge results shown in (e) are identical to those shown in Figures 6(i), 7(i), and 9(h).

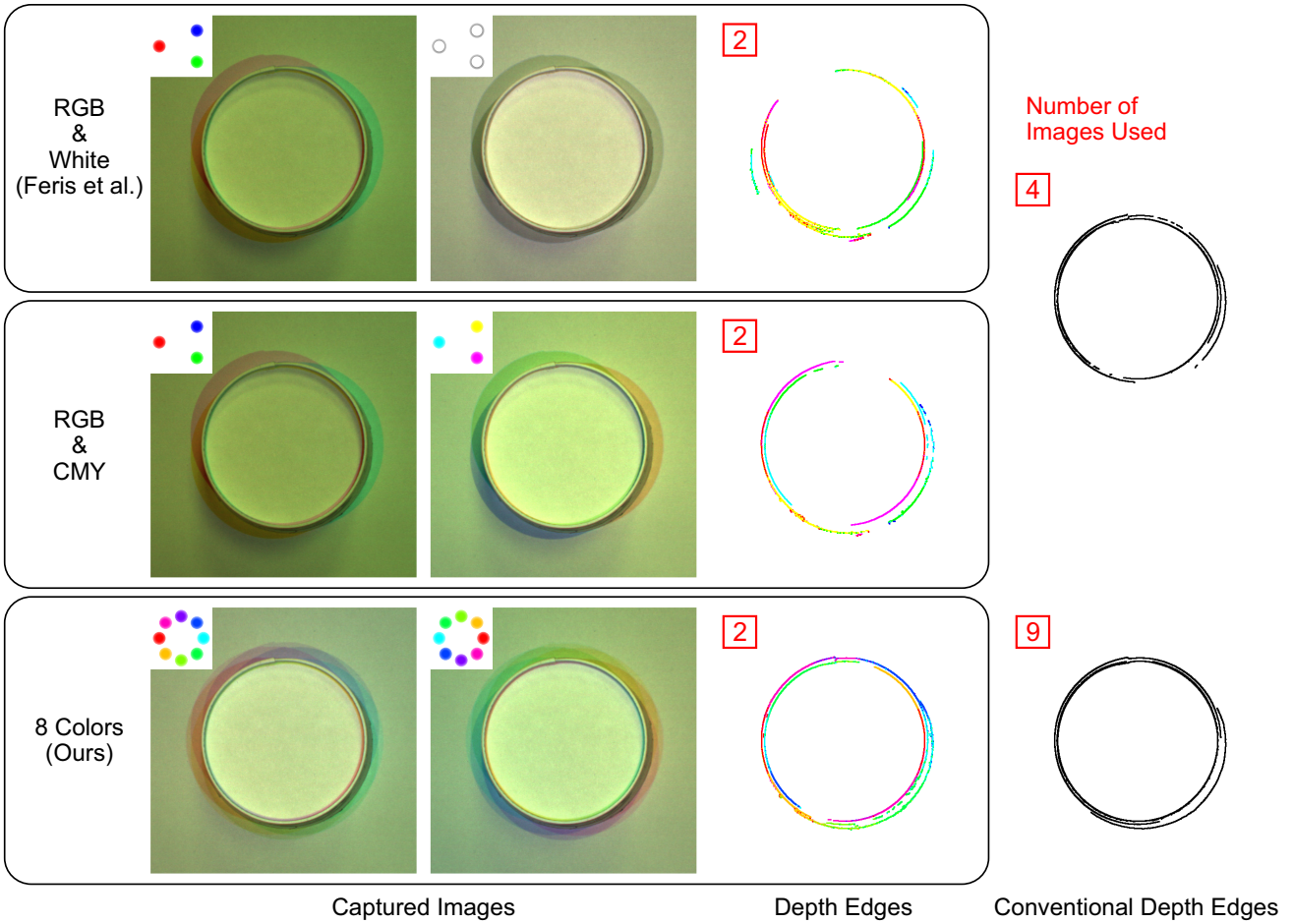

Fig. 12 Comparison with Feris et al. [8] for a scene including the bottom of a paper cup. (Top) The depth edges computed by Feris et al.'s method are noisy and include incorrect orientations due to the limited color information available only in the first image. (Middle) Instead of using the white light sources, we can capture the second image using the complementary colors (CMY) to make the distance measure more robust, producing better depth edges. (Bottom) Our method allows any number of light sources multiplexed. Using 8 light sources further improves the depth edges. Conventional depth edges extracted using 3 and 8 light sources are also shown as references.

and M_2. We used as A the ambient images captured for the conventional MFC method. We generated W by simulating the illumination using the conventional MFC images as $W = \frac{1}{t}(\sum I_i^s + A)$, where t is a scale factor[3].

As can be seen from the figure, our shadow confidence map $D_{\mathrm{HS}}(M_1, M_2)$ provides the highest contrast between shadow and non-shadow regions, because the colors of shadow regions become complementary in the two images, while the colors of non-shadow regions become similar in the two images. Generally in our experimental setup, A is too dark to correctly compute the hue and saturation in both shadow and non-shadow regions; thus $D_{\mathrm{HS}}(M_1, A)$ does not separate the two regions well as shown in Figure 11(a). W is captured with neutral (white) color illuminations corresponding to the center of the hue-saturation plane; thus $D_{\mathrm{HS}}(M_1, W)$ leads to a lower contrast than $D_{\mathrm{HS}}(M_1, M_2)$ as shown in Figure 11(b). Note also that W does not convey any information on the shadow colors. Therefore the shadow color check process in Step 6 of our algorithm can use only the first image M_1. The low contrast shadow confidence map as well as the single-image color check process due to the use of W as the second image result in more false depth edges shown in Figure 11(d). Our depth edges obtained using M_1 and M_2 are much clearer as shown in Figure 11(e).

We also tested a three-shot approach, where we used M_1, M_2, and A, and computed the shadow confidence map as $D_{\mathrm{HS}}(M_1 - A, M_2 - A)$. This three-shot approach produced very similar results as our approach using $D_{\mathrm{HS}}(M_1, M_2)$, because (1) ambient lights (as well as scene textures) are canceled by computing $D_{\mathrm{HS}}(M_1, M_2)$, and (2) our light sources are substantially stronger than ambient lights in our experiments as described in Section 3.1. Such a three-shot approach might add some robustness when we capture images with stronger ambient lights, although it increases the

[3] We observed that $D_{\mathrm{HS}}(M_1, W)$ is not sensitive to the scale factor t as long as the pixels in W are not saturated and not too dark. This is expected because t only affects the brightness (value) component of the image. In experiments we set $t = 4$.

Table 1 Performance comparisons of the bin-picking systems using the conventional MFC [12] and our rainbow flash camera. The normalized cycle time is computed by dividing the cycle time by the pickup success rate, which measures the average time required for a single pickup success.

	Pickup Performance				Time (seconds)		
	Total Trial	Success	Failure	Success Rate	Capture Time (2 Views)	Cycle Time	Normalized Cycle Time
Conventional	303	283	20	93.4%	1.2	8	8.6
Ours	302	274	28	90.7%	0.26	7	7.7

capture time. If the ambient lights are much stronger than the light sources (e.g., under outdoor lighting conditions), even such an approach would not work, which is the case in most of active illumination systems.

4.3 Comparisons with the RGB Multiplexing Method

Feris et al. [8] captured two images for general scenes: one with three red, green, and blue light sources on and the other by replacing them with white light sources. We realized it by assigning the colors to three light sources at the positions of $\{0^\circ, 135^\circ, 225^\circ\}$ as shown in the insets of Figure 12. We computed a shadow confidence map by taking the ratio between each channel of the two intensity-normalized images as described in [8], and generated depth edges by applying Steps 4 to 8 in our method. We used only the first image in the shadow color check process (Step 6), since the second image does not convey color information.

If the three light sources were positioned at $\{0^\circ, 120^\circ, 240^\circ\}$, Feris et al.'s method would be a special case of our method, capturing the first image with a hue circle approximated by three light sources and the second image with white light sources. However, as shown in Section 4.2, it provides a less reliable distance measure than ours, resulting in noisy depth edges as shown in Figure 12 (top). Moreover, because of the limited color information available only in the first image, orientations of some depth edges are misclassified. If we used three light sources, we would capture the second image with complementary colors of RGB (i.e., CMY) to make the distance measure more robust, producing better depth edges as shown in Figure 12 (middle). While the standard RGB multiplexing is applicable only to three light sources, our method allows any number of light sources multiplexed to approximate a hue circle. Our method using 8 light sources further improves the depth edges as shown in Figure 12 (bottom).

4.4 Bin Picking Using the Rainbow Flash Camera

We built a bin-picking system using the rainbow flash camera on Liu et al.'s system [12]. Liu et al. used a conventional MFC mounted on a robot arm to extract depth edges of objects randomly placed in a bin, estimated the pose of an object using the FDCM algorithm followed by refinement based on the iterative-closest point (ICP) algorithm, and grasped the object using the estimated pose. To increase the pose estimation accuracy, a single grasping cycle included two MFC captures from different viewpoints, which were jointly used in the pose refinement process. We replaced the conventional MFC with our rainbow flash camera, while using the same system pipeline and pose estimation algorithms. We used the objects shown in Figure 9 as the target. We used the two-shot algorithm for robust depth edge extraction.

The supplementary video and Table 1 compare the cycle times of the conventional and our bin-picking systems. Since the frame rate of the camera was at 15 frames per second, the capture time at a single viewpoint was reduced from 0.6 seconds (9 images) to 0.13 seconds (2 images), resulting in a total of 0.93 seconds reduction for the two viewpoints in a single cycle. The average cycle time was reduced from about 8 seconds to 7 seconds as shown in the video. Note that the pose estimation process requires less than 1 second and the computation occurs during the robot motion [12]. Thus, most of the cycle time is taken by the robot motion in our system using the rainbow flash camera, because the system requires only 0.26 seconds to capture images without any robot motion. Different system designs that require a shorter time for robot motion would further reduce the cycle time.

We also measured the pickup success rates of the conventional and our bin-picking systems using a similar setting to [12]. Table 1 summarizes the results. We observed that the pickup success rate of our system (90.7%) was slightly worse than that of the conventional system (93.4%). In particular, since this specific object has similar depth edges when it is flipped upside-down [12], missing and noisy depth edges caused by the color multiplexing of our method resulted in a larger number of incorrect estimation corresponding to a flipped pose. Those pickup errors could be detected by using some error detection mechanism (e.g., by using another camera looking at the gripper [12]) after each pickup trial. Given such an error detection mechanism, the average time required for a single pickup success can be measured by the cycle time normalized by the pickup success rate. Table 1 clarifies that our system has the advantage over the conventional system even in terms of the normalized cycle time (after we take the pickup success rate into account).

5 Conclusions

We presented a novel color multiplexing method for depth edge extraction that exploits the complementary nature of a hue circle. Our method colorizes shadow regions in captured images because of the occlusion of a half of the colors in the hue circle, while letting non-shadow regions have a neutral color because of the mixture of all the colors in the hue circle. We presented a single-shot depth edge extraction algorithm that analyzes the colored shadows in a single captured image. Although the single-shot algorithm produces better depth edges than standard intensity-based edge detectors, it is still unstable in uncontrolled environments. To robustly handle scene textures and non-Lambertian reflections, we presented a depth edge extraction algorithm using two images captured with a hue circle and its complementary version. We showed a prototype of the rainbow flash camera using 8 color LEDs and discussed the advantages and limitations of our approach compared to conventional methods under various scenes. We also demonstrated a practical bin-picking system using the rainbow flash camera mounted on a robot arm, which reduces the capture and cycle time of the system using the conventional MFC while maintaining a pickup success ratio of more than 90%.

Acknowledgements The author thanks Jay Thornton for many valuable discussions and for naming the proposed camera. The author also thanks Ramesh Raskar, Amit Agrawal, Oncel Tuzel, Srikumar Ramalingam, Tim K. Marks, Ming-Yu Liu, Makito Seki, and Yukiyasu Domae for their feedback and support, and the anonymous reviewers of ECCV 2012 and this journal submission for their helpful comments. Special thanks to John Barnwell, William Yerazunis, and Abraham Goldsmith for their help in developing the camera prototype. The final publication is available at Springer via `http://dx.doi.org/10.1007/s11263-014-0726-4`.

References

1. Agrawal, A., Sun, Y., Barnwell, J., Raskar, R.: Vision-guided robot system for picking objects by casting shadows. Int'l J. Robotics Research **29**(2–3), 155–173 (2010)
2. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
3. Chen, C., Vaquero, D., Turk, M.: Illumination demultiplexing from a single image. In: Proc. IEEE Int'l Conf. Computer Vision (ICCV) (2011)
4. Crispell, D., Lanman, D., Sibley, P.G., Zhao, Y., Taubin, G.: Beyond silhouettes: Surface reconstruction using multi-flash photography. In: Proc. Int'l Symp. 3D Data Processing, Visualization, and Transmission (3DPVT), pp. 405–412 (2006)
5. De Decker, B., Kautz, J., Mertens, T., Bekaert, P.: Capturing multiple illumination conditions using time and color multiplexing. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 2536–2543 (2009)
6. Feris, R., Raskar, R., Chen, L., Tan, K.H., Turk, M.: Multiflash stereopsis: Depth-edge-preserving stereo with small baseline illumination. IEEE Trans. Pattern Anal. Mach. Intell. **30**(1), 147–159 (2008)
7. Feris, R., Raskar, R., Tan, K.H., Turk, M.: Specular reflection reduction with multi-flash imaging. In: Proc. Brazilian Symp. Computer Graphics and Image Processing (SIBGRAPI), pp. 316–321 (2004)
8. Feris, R., Turk, M., Raskar, R.: Dealing with multi-scale depth changes and motion in depth edge detection. In: Proc. Brazilian Symp. Computer Graphics and Image Processing (SIBGRAPI), pp. 3–10 (2006)
9. Fyffe, G., Yu, X., Debevec, P.: Single-shot photometric stereo by spectral multiplexing. In: Proc. IEEE Int'l Conf. Computational Photography (ICCP), pp. 1–6 (2011)
10. Hernandez, C., Vogiatzis, G., Brostow, G.J., Stenger, B., Cipolla, R.: Non-rigid photometric stereo with colored lights. In: Proc. IEEE Int'l Conf. Computer Vision (ICCV), pp. 1–8 (2007)
11. Liu, M.Y., Tuzel, O., Taguchi, Y.: Joint geodesic upsampling of depth images. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2013)
12. Liu, M.Y., Tuzel, O., Veeraraghavan, A., Taguchi, Y., Marks, T.K., Chellappa, R.: Fast object localization and pose estimation in heavy clutter for robotic bin picking. Int'l J. Robotics Research **31**(8), 951–973 (2012)
13. MacEvoy, B.: Color vision. `http://www.handprint.com/LS/CVS/color.html` (2008)
14. Minomo, Y., Kakehi, Y., Iida, M., Naemura, T.: Transforming your shadow into colorful visual media: Multiprojection of complementary colors. Comput. Entertain. **4**(3) (2006)
15. Park, J.I., Lee, M.H., Grossberg, M.D., Nayar, S.K.: Multispectral imaging using multiplexed illumination. In: Proc. IEEE Int'l Conf. Computer Vision (ICCV) (2007)
16. Raskar, R., Tan, K.H., Feris, R., Yu, J., Turk, M.: Non-photorealistic camera: Depth edge detection and stylized rendering using multi-flash imaging. ACM Trans. Graphics **23**(3), 679–688 (2004)
17. Sá, A.M., Carvalho, P.C.P., Velho, L.: (b, s)-BCSL: Structured light color boundary coding for 3D photography. In: Proc. Vision, Modeling, and Visualization Conf. (VMV) (2002)
18. Schechner, Y., Nayar, S., Belhumeur, P.: A theory of multiplexed illumination. In: Proc. IEEE Int'l Conf. Computer Vision (ICCV), vol. 2, pp. 808–815 (2003)
19. Shroff, N., Taguchi, Y., Tuzel, O., Veeraraghavan, A., Ramalingam, S., Okuda, H.: Finding a needle in a specular haystack. In: Proc. IEEE Int'l Conf. Robotics Automation (ICRA), pp. 5963–5970 (2011)
20. Taguchi, Y.: Rainbow flash camera: Depth edge extraction using complementary colors. In: Proc. European Conf. Computer Vision (ECCV), vol. 6, pp. 513–527 (2012)
21. Vaquero, D.A., Feris, R.S., Turk, M., Raskar, R.: Characterizing the shadow space of camera-light pairs. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (2008)
22. Vaquero, D.A., Raskar, R., Feris, R.S., Turk, M.: A projector-camera setup for geometry-invariant frequency demultiplexing. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), pp. 2082–2089 (2009)
23. Wan, G., Horowitz, M., Levoy, M.: Applications of multi-bucket sensors to computational photography. Tech. Rep. 2012-2, Stanford Computer Graphics Laboratory (2012)
24. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Opt. Eng. **19**(1), 139–144 (1980)