Diffuse-Specular Separation and Depth Recovery from Image Sequences

Stephen Lin[1], Yuanzhen Li[1,2], Sing Bing Kang[3], Xin Tong[1], and
Heung-Yeung Shum[1]

[1] Microsoft Research, Asia
[2] Chinese Academy of Science [†]
[3] Microsoft Research

Abstract. Specular reflections present difficulties for many areas of
computer vision such as stereo and segmentation. To separate specu-
lar and diffuse reflection components, previous approaches generally re-
quire accurate segmentation, regionally uniform reflectance or structured
lighting. To overcome these limiting assumptions, we propose a method
based on color analysis and multibaseline stereo that simultaneously es-
timates the separation and the true depth of specular reflections. First,
pixels with a specular component are detected by a novel form of color
histogram differencing that utilizes the epipolar constraint. This process
uses relevant data from all the stereo images for robustness, and ad-
dresses the problem of color occlusions. Based on the Lambertian model
of diffuse reflectance, stereo correspondence is then employed to compute
for specular pixels their corresponding diffuse components in other views.
The results of color-based detection aid the stereo correspondence, which
determines both separation and true depth of specular pixels. Our ap-
proach integrates color analysis and multibaseline stereo in a synergistic
manner to yield accurate separation and depth, as demonstrated by our
results on synthetic and real image sequences.

1 Introduction

The difference in behavior between diffuse and specular reflections poses a prob-
lem in many computer vision algorithms. While purely diffuse reflections ordi-
narily exhibit little variation in color from different viewing directions, specular
reflections tend to change significantly in both color and position. A vast ma-
jority of techniques in areas such as stereo and image segmentation account
only for diffuse reflection and disregard specularities as noise. While this diffuse
assumption is generally valid for much of an image, processing of regions that
contain specular reflections can result in significant inaccuracies. For instance,
traditional stereo correspondence of specular reflections does not give the true
depth of a scene point, but instead an incorrect virtual depth. To avoid such
problems, methods for image preprocessing have been developed to detect and
to separate specular reflections in images.

[†] This work was performed while the second author was visiting Microsoft Research,
Asia.

1.1 Related Work

To distinguish between diffuse and specular image intensities, most separation methods utilize color images and the dichromatic reflectance model proposed by Shafer [15]. This model suggests that, in the case of dielectrics (non-conductors), diffuse and specular components have different spectral distributions. The spectral distribution of the specular component is similar to that of the illumination, while the distribution of the diffuse component is a product of illumination and surface pigments. The color of a given pixel can be viewed in RGB color space as a linear combination of a vector for object reflectance color and a vector for illumination color. All image points on a uniform-colored surface lie on a dichromatic plane which is spanned by these two vectors.

Rich literature exists on separation using the dichromatic model. Klinker *et al.* [7] developed a method based on the observation that the color histogram of a surface with uniform reflectance takes the shape of a "skewed T" with two limbs. One limb represents purely diffuse points while the other corresponds to highlight points. Their separation algorithm automatically identifies the two limbs, then computes the diffuse component of each highlight point as its projection on the diffuse limb. To estimate specular color, Tong and Funt [16] suggest computing the dichromatic planes of several uniform-reflectance regions and finding the line most parallel to these planes. Sato and Ikeuchi [14] also employed the dichromatic model for separation by analyzing color signatures produced from many images taken under a moving light source.

Besides color, polarization has also been an effective cue for specular separation. Wolff and Boult [17] proposed a polarization-based method for separating reflection components in regions of constant Fresnel reflectance. Nayar *et al.* [12] used polarization in conjunction with color information from a single view to separate reflection components, where a constraint is provided by neighboring pixels with similar diffuse color.

These previous approaches have produced good separation results, but they have requirements that limit their applicability. Nearly all of these methods assume some sort of consistency among neighboring pixels. For the color algorithms, color segmentation is needed to avoid problematic overlaps in color histograms among regions with different diffuse reflection color. Segmentation algorithms are especially unreliable in regions containing specularity, and errors in segmentation significantly degrade the results of dichromatic plane analysis. Additionally, these color methods generally assume uniform illumination color, so interreflections cannot be present. For typical real scenes, which have textured objects and interreflections, anatomizing the color histogram is essentially infeasible. The need for segmentation also exists for polarization methods, which additionally require polarizer rotation. The algorithm by Sato and Ikeuchi [14] notably avoids these assumptions, but the need for an image sequence taken under a moving light source restricts its use.

1.2 Stereo in the Presence of Specularities

Stereo in the presence of specular reflection has been a challenging problem. Bhat and Nayar [1] consider the likelihood of correct stereo matching by analyzing

the relationship between stereo vergence and surface roughness, and in [2] they further propose a trinocular system where only two images are used at a time in the computation of depth at a point. Jin *et al.* [5] poses this problem within a variational framework and seeks to estimate a suitable shape model of the scene. Of these methods, only [2] has endeavored to recover true depth information for specular points, but it involves extra efforts to determine a suitable trinocular configuration.

1.3   Our Approach: Color-Based Separation with Multibaseline Stereo

We circumvent the limiting assumptions made in previous works by framing the separation problem as one of stereo correspondence. By matching specular pixels to their corresponding diffuse points in other views, we can determine the diffuse components of the specularities using the Lambertian model of diffuse reflection. In this way, we employ multibaseline stereo in a manner that computes separation and depth together.

Without special consideration of camera positions, we handle highlights in multibaseline stereo by forming correspondence constraints based on the following assumptions: diffuse reflection satisfies the Lambertian property; specular reflections vary in color from view to view; scene points having specular reflection exhibit purely diffuse reflection in some other views. The first assumption is commonplace in computer vision, and the second assumption is not unreasonable, because specular reflections change positions in the scene for different viewpoints. Even when a specular reflection lies on the same uniformly colored surface between two views, the underlying diffuse shading and the specular intensity will likely change, thus yielding different color measurements [9]. The third assumption is dependent on surface roughnesses in the scene, but is often satisfied by having a sufficiently long stereo baseline for the entire image sequence.

With epipolar geometry and the aforementioned viewing conditions, our proposed method first detects specular reflections using stereo-enhanced color processing. The color-based detection results are then used to heighten the accuracy of stereo correspondence, which yields simultaneous separation and depth recovery. It is this synergy between color and stereo that leads to better estimation of both separation and depth. With their combination, we develop an algorithm that is novel in the following aspects:

- The epipolar constraint is used to promote color histogram differencing (CHD) by correspondence of image rows.
- A technique we call *Multi-CHD* is introduced to handle color occlusions and to make full use of available images for robust specularity detection.
- Diffuse components and depth of specular reflections are computed by stereo correspondence constrained by epipolar geometry and detection results.

In comparison with previous separation approaches, our use of multibaseline stereo provides significant practical advantages. First, prior image segmentation and consistency among neighboring pixels are not assumed. Second, our method

is robust to moderate amounts of interreflection, which are present in most real scenes. Third, capture of image sets can potentially be done instantaneously, since we do not require polarizer rotations or changes in lighting conditions that prohibit possible frame-rate execution. These three properties make our method more feasible than any previous method.

The remainder of this paper is organized as follows. Section 2 describes the details of specularity detection by color histogram differencing under epipolar constraints. These detection results are used in Section 3 to constrain stereo correspondence, which determines specularity separation and depth. The effectiveness of our algorithm is supported by experimental results presented in Section 4, and the paper closes with a discussion and some conclusions in Sections 5 and 6, respectively.

2   Color Histogram Differencing

For specularity detection, we develop a novel method that takes advantage of multibaseline stereo to improve color histogram differencing. CHD, introduced in [9], is based upon changes in the color of a specular reflection from view to view. Under the Lambertian property, the diffuse colors of a scene are viewpoint independent, with the exception of occlusions and disocclusions. Since the scene location of a specular reflection changes from view to view, its underlying diffuse reflection can differ in shading or color. The underlying diffuse color is a component of the overall specular color, so a specular reflection will change in color between different views.

This difference in behavior between specular and purely diffuse reflections is exploited by CHD to detect specularities. Suppose that we have two images I_1 and I_2 of the same scene taken from different viewpoints. For each image I_k, its pixel colors can be mapped into a binary RGB histogram H_k, such as those shown in Fig. 1. If a scene point contains purely diffuse reflection, then its positions in H_1 and H_2 will be identical. A specular reflection in H_1, however, will have a shifted histogram position in H_2. By subtracting points in H_2 from H_1, we can detect the specular pixels in I_1:

$$H_{1,spec} = H_1 - H_2 = \{p \mid p \in H_1, \ p \notin H_2\}.$$

Because of image noise and other unmodelled effects, some leeway is allowed in the differencing such that slight differences in histogram point positions between H_1 and H_2 do not indicate specularity.

In the remainder of this section, we present our method of specularity detection based on CHD. First we describe tri-view CHD, which extends standard CHD to a trinocular multibaseline stereo setting in a way that takes advantage of epipolar geometry and addresses color occlusion. Then we generalize tri-view CHD to make full use of the abundant stereo images in a method we call *multi-CHD*.
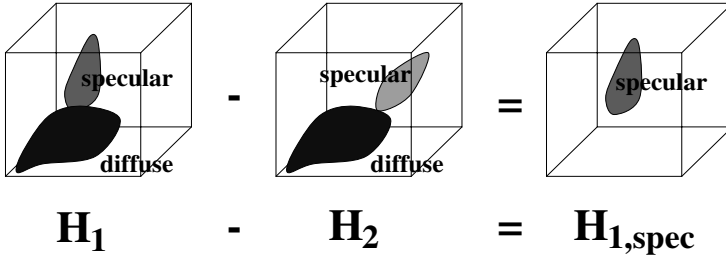
Fig. 1. Standard CHD. Since specular reflections generally change in color from one view to another, they can be detected by histogram differencing.

2.1   Tri-View CHD

Two major obstacles for standard CHD are histogram clutter and color occlusions. When parts of a histogram are crowded with points, CHD can fail because specular histogram points in one view can possibly be masked by other points, diffuse or specular, with similar color in another view. When this happens, specular pixels go undetected. The change in viewpoint also presents difficulties for CHD because diffuse colors present in one view may be occluded in another view by scene geometry or specular reflections. As a result, diffuse pixels are mistakenly detected as specular. To address both of these problems, we present a method called *tri-view CHD*.

In standard CHD, differencing is done between histograms formed from entire images. Because of the large number of pixels involved, there often exists much crowding in color histograms. To reduce this clutter, we take advantage of epipolar geometry, which allows us to decrease the number of histogram points by differencing image rows rather than entire images. In multibaseline stereo, the correspondence between images rows is simply the matching scanlines. For tri-view CHD, matching scanlines for three images are illustrated in Fig. 2 and are denoted as I_C for the central (reference) viewpoint, I_L for the left viewpoint, and I_R for the right viewpoint. The histograms corresponding to these scanlines are respectively labelled H_C, H_L and H_R.

We use three images to lessen the effects of occlusion caused by both geometry and specular reflections. Diffuse reflection in I_C that is geometrically occluded in I_L or I_R are present in the histogram $H_L \cup H_R$, because different parts of the scene are occluded when changing the viewpoint to the left or to the right. The same is true for diffuse reflection in I_C that is occluded by specularity in I_L or I_R. Based on this observation, we can formulate the set of histogram points in H_C that contain specular reflection as

$$H_{C,spec} = H_C - (H_L \cup H_R).$$

The detected specular histogram points are backprojected to the image to locate the specular pixels, which we represent as a binary image:

$$S_C(x,y) = \begin{cases} 0 & \text{if } I_C(x,y) \text{ is non-specular} \\ 1 & \text{if } I_C(x,y) \text{ is specular.} \end{cases} \tag{1}$$
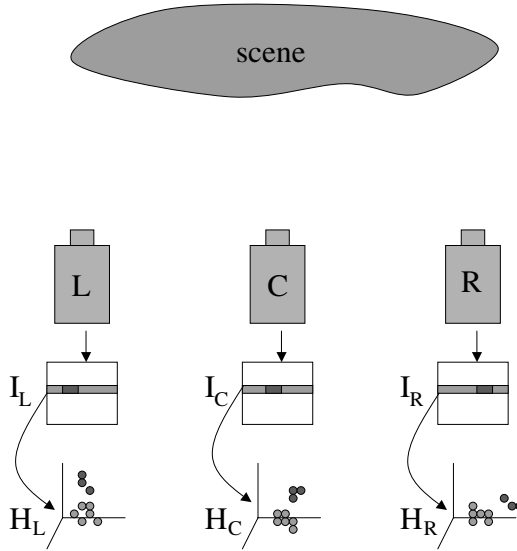
Fig. 2. Tri-view CHD. In the multibaseline camera configuration, matching epipolar lines are used for histogram differencing.

This consideration of occlusion effects and cluttering in CHD diminishes both the number of false positives and the number of missed detections.

Although we can compute the detection using tri-view CHD, the results can be sensitive to viewpoint intervals. For intervals that are too small, specularity color may not differ significantly, and for very large intervals visibility differences arise. Since an appropriate interval is dependent on scene geometry and surface roughnesses, it is impractical to compute. But since a multibaseline stereo sequence contains many viewpoints from which to form image triplets, we can make use of these available images for more robust detection.

2.2   Multi-CHD

We refer to use of the entire stereo sequence for tri-view CHD as *multi-CHD*. To explain this technique, let us consider the camera configuration illustrated in Fig. 3, with $(2n+1)$ viewpoints from which we capture $(2n+1)$ images labelled as $\{I_{L_n} \ldots I_{L_2}, I_{L_1}, I_C, I_{R_1}, I_{R_2} \ldots I_{R_n}\}$. For the center reference image I_C, we can form n different uniform-interval triplets $\{I_{L_k}, I_C, I_{R_k}\}$ for $k = 1, 2 \ldots n$. Tri-view CHD can be performed on each of these triplets as described in the previous subsection to form k specular point sets $S_{C,k}$. These sets are used to vote for the final detection results.

Voting is an effective method for obtaining a more reliable result by combining data from a number of sources. The voted output reflects a consensus or compromise of the information. For each image pixel, it is deemed specular if the number of votes from $S_{C,k}$ exceeds a given threshold t, i.e., if $\sum_{k=1}^{n} S_{C,k} > t$.
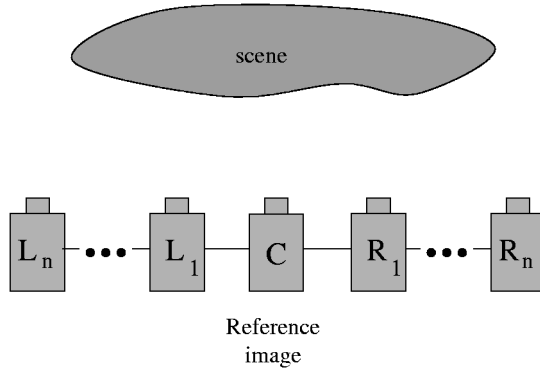
Fig. 3. Camera configuration for Multi-CHD.

Although we employ this basic voting scheme, it may be enhanced by weighting votes according to its distance from the reference viewpoint. Additionally, the reliability of voting can be improved by increasing the number of statistically independent voters [8]. This can be done by considering image triplets where the viewpoints are unevenly spaced, or by generalizing the multi-CHD concept to a 2D grid of cameras.

3   Stereo Correspondence for Specular Reflections

Once we detect the specular pixels in an image, we attempt to associate them to their corresponding diffuse points in other images to determine separation and depth. Because of the large differences in intensity and color, a correspondence cannot directly be computed between a specular pixel and its diffuse counterpart, so we instead correspond diffuse pixels of other images under the constraint of their disparity relationship to the specular pixel in the reference image.

In our approach, we first use multibaseline stereo to compute initial disparity estimates for diffuse reflection. With these diffuse disparities, we impose a continuity constraint to determine a search range for estimation of specular disparities. This second round of stereo combines the use of spatially flexible windows and a dynamically selected subset of images with which to compute the matches. Upon determining the correspondence, a diffuse component is estimated for producing the separation, and the computed disparity gives the true depth.

3.1   Initial Depth Estimation for Diffuse Reflection

Our specular correspondence method begins with an initial depth estimate that employs the epipolar constraint. This preliminary estimate is reasonably accurate for purely diffuse pixels, and it is used to facilitate more precise depth

estimation for specular pixels. The uncertainty of this estimate can also be used as an additional independent cue for specularity detection.

For a forward-facing multibaseline stereo configuration [13], disparity varies linearly with horizontal pixel displacement. In order to estimate the disparity for a pixel (x, y), we first aggregate the matching costs over a window as the sum of sum of squared differences (SSSD), namely

$$E_{SSSD}(x, y, d) = \sum_{k \neq 0} \sum_{(u,v) \in W(x,y)} \rho\big(I_0(u, v) - \hat{I}_k(u, v, d)\big), \qquad (2)$$

where $\rho(\bullet)$ is the per-pixel squared Euclidean distance in RGB between reference image I_0 and $\hat{I}_k$ (warped image of I_k at disparity d). $W(x, y)$ is a square window centered at (x, y).

For each non-specular pixel (x, y) in the reference image as determined by multi-CHD, the minimum E_{SSSD} for different disparity values determines the estimated disparity d and the uncertainty u of the estimation:

$$d(x, y) = arg \min_d E_{SSSD(x,y,d)} \quad \text{and} \quad u(x, y) = \min_d E_{SSSD(x,y,d)}.$$

Similar to confidence measures in other stereo works, u gauges uncertainty because match quality is poor when its value is high. This quantity will later be used within a confidence weight for disparity estimates, and additionally, it can be used as a cue for specular reflection. Pixels with uncertainty above a specified threshold can be added to the multi-CHD detection results.

To handle occlusions which pose a problem in dense multi-view stereo, we incorporate shiftable windows and temporal image selection [6] to improve the matching of boundary pixels and semi-occluded pixels. Since specular points cannot be matched in this way, we more carefully correspond them using the continuity constraint.

3.2   Continuity Constraint

With the computed disparity estimates for diffuse pixels, we calculate a disparity search interval for specular pixels using the continuity constraint. Based on the cohesiveness of matter, the continuity constraint proposed by Marr and Poggio [10] claims that disparity should vary smoothly throughout an image, given opaque material and piecewise continuous surfaces. The failure of this assumption at depth discontinuities can be remedied by using a large window for SSSD computation. We can safely impose this constraint to get an initial estimate and search range of disparity for specular points, even if they are located near depth discontinuities.

We implement the continuity constraint using k-Nearest Neighbors (k-NN). For a particular specular point $S(x, y)$, we find k nearest neighbor points that are nonspecular around it. Their disparity estimates d_l for $l = 1 \ldots k$ are accompanied by corresponding uncertainty values u_l that indicate estimation quality. The initial estimate for disparity is expressed in terms of the k nearest neighbor disparities weighted by the reciprocal of their uncertainty values, i.e., $d_i(x, y) = \sum_l d_l u_l^{-1} / \sum_l u_l^{-1}$.

From this initial estimate, we restrict the search range of possible disparities to be within a preset bound d_r so that the disparity candidate interval $D(x,y)$ for $S(x,y)$ is

$$D(x,y) \ = \ \big[\ d_i(x,y) - d_r\ ,\ d_i(x,y) + d_r\ \big]. \tag{3}$$

We take d_r in our experiments to be equal to the standard deviation of the initial disparity estimates over the image.

3.3   Shiftable and Flexible Windows

Since pixels with specular reflection degrade area-based correlation, we employ shiftable and flexible windows, rather than a fixed window, to get a selectively aggregated matching error. The basic idea of shiftable windows is to examine several windows that include the pixel of interest, not just the window centered at that pixel. This strategy has been shown to effectively deal with occlusions [11] [3]. We extend this idea by excluding pixels detected as specular in the previous section. This results in shiftable windows with flexible shapes, and we show it improves the matching of pixels that are specular in some images and non-specular in others. This is furthermore effective in dealing with pixels near the boundary between specular and diffuse regions.

In the formation of these flexible windows, we use the specular detection S_k from (1) for respective images I_k. For correspondence between image pairs I_0 and I_k, we modify the $n \times n$ shiftable window $W_{n \times n}(x,y)$ that includes (x,y) into the window $W_f(x,y)$ whose support is flexibly shaped to exclude specularity:

$$W_f(x,y) = \big\{(u,v) | (u,v) \in W_{n \times n}(x,y), S_0(u,v) = 0, S_k(u,v) = 0\big\}. \tag{4}$$

In the case where the number of valid pixels in a window falls below a threshold, we progressively increase the original window size.

Over this flexible and shiftable window, we aggregate the raw matching cost to compute the SSD:

$$E_{SSD}(x,y,d,k) = \frac{\sum\limits_{(u,v) \in W_f(x,y)} w(u,v) E_{raw}(u,v,d,k)}{\sum\limits_{(u,v) \in W_f(x,y)} w(u,v)},$$

where $w(u,v)$ is the support weight of each pixel in $W_f(x,y)$ for (x,y), which we set to the constant 1 to get the mean.

3.4   Temporal Selection

Rather than summing the match costs over all views, a better approach would be to dynamically select a subset of views where the support window is believed to be mostly diffuse and unoccluded. Towards this end, we can formulate from the specular detection results a temporally selective aggregated matching error:

$$E_{SSSD}(x,y,d) = \frac{\sum\limits_{k \neq 0\ ,\ C(x,y)>T} wt(k) E_{SSD}(x,y,d,k)}{\sum\limits_{k \neq 0\ ,\ C(x,y)>T} wt(k)},$$

where

$$C(x, y) = \sum_{(u,v) \in W_f(x,y)} [1 - S_k(u, v)]. \tag{5}$$

The constraint $C(x, y) > T$ ensures that in the selected views the correlation window includes an appropriate number of diffuse points, where T is a percentage of pixels in the original $n \times n$ shiftable window. The factors $wt(k)$ are weights of $E_{SSD}(x, y, d, k)$ which could normalize for the number of temporally selected views. We instead use these weights to deal with occlusions in the selected views. Views with a lower local SSD error $E_{SSD}(x, y, d, k)$ are more likely to have visible corresponding pixels, so we set $wt(k) = 1$ for the best 50% of images satisfying constraint (5), and $wt(k) = 0$ for the remaining 50%. This temporal selection rule is similar to that described in [6].

Finally, we adopt a winner selection strategy to compute the final disparity:

$$d(x, y) = arg \min_{d \in D(x,y)} E_{SSSD}(x, y, d),$$

where $D(x, y)$ is the candidate interval for disparity given in (3).

3.5 Separation and Depth Estimate

Upon corresponding a specular point P_S to a diffuse point P_D in any other view, we can directly compute depth from the disparity and can theoretically take the color of P_D as the diffuse component of P_S, because of the Lambertian diffuse reflection assumption. The specular component can simply be computed as the difference in color between P_S and P_D. In reality, we must also account for noise and mismatches. So we find all possible corresponding diffuse points in the light field for P_S, and then compute the mean value of the color measurements to obtain P_D.

4 Experimental Results

In this section, we present results on synthetic and real image sequences to validate our approach. For both sequences, our algorithm settings are a multi-CHD voting threshold of 50% and a CHD differencing distance threshold equal to one standard deviation of the image noise. The depth interval values d_4 in (3) are 1.4 for synthetic images and 2.2 for real images.

4.1 Synthetic Image Sequence

Experiments on synthetic images are used for comparison to ground truth data. Our synthetic sequence consists of 57 images generated using Phong shading. The baseline distance between consecutive views is 3.125mm.

Fig. 4 displays images of our results. An image taken from our sequence is shown in (a). For this image, ground truth and our own detection images are exhibited in (b) and (c) respectively. When the scene patch for a pixel contains more than one color, color blending occurs because of pixel integration. The

balance of this blending along color boundaries will change from view to view, so these pixels are mistakenly detected as specular by CHD.

In (d-f), an image comparison of depth recovery is shown. Our depth estimation (f) more closely resembles the ground truth (d), because of our consideration of specular reflections. When specularities are disregarded as in (e), incorrect virtual depths are computed instead.

The images (g-j) show the ground truth separation and our own. Our result is similar to the ground truth, even though several pixels of the fruit in the bowl contain specularity throughout our image sequence. When this happens, our method tends to match with neighboring pixels, which gives a reasonable approximation of the diffuse component.

4.2   Real Image Sequence

For a real sequence, we use 64 images with a baseline distance between consecutive views of 7.8125mm and a focal length of 60mm. For SSD matching, we use a window size of 9x9 and only eleven consecutive images for depth recovery, to avoid non-constant disparities that can result from large baselines. Fig. 5(a) displays an image from this sequence, and (b) exhibits our detection result by multi-CHD. There exist some false detections mainly due to color blending, but the true specular areas are mostly found. For specular reflections on the wooden bowl, the watermelon and the leaves, the depth recovery without consideration of specular reflections in (c) gives an inaccurate virtual depth, while our estimation in (d) appears to approximate the true depth. The apple presents difficulties due to lack of texture, which is problematic for all stereo methods. Our diffuse and specular separation results are shown in (e) and (f), respectively. Notice that although depth recovery is inaccurate for textureless regions like the apple, its separation result is not bad because correspondence is made with some other points in the textureless region, whose diffuse component is often close to being correct.

Differences between tri-view CHD and standard (full-image) CHD are demonstrated in Fig. 6. In (a), we show the result of standard CHD between the reference image and the image taken 7.8125cm to the left. Because of color occlusions due to differences in scene visibililty, standard CHD detects many false positives on the right side of the image. Moreover, for the large specular area on the wooden bowl, there are many missing detections that result from histogram crowding. For standard CHD with the image taken 7.8125cm to the right, it is seen in (b) that color occlusions cause many false positives on the left side, and there are missing detections on both the wooden bowl and the watermelon. Tri-view CHD using these three images produces the result in (c), in which the color occlusion and histogram crowding problem is better handled. A combination of (a) and (b) cannot generate the results in (c), and we improve upon this result using all views in multi-CHD to get the detection shown in Fig 5(b).
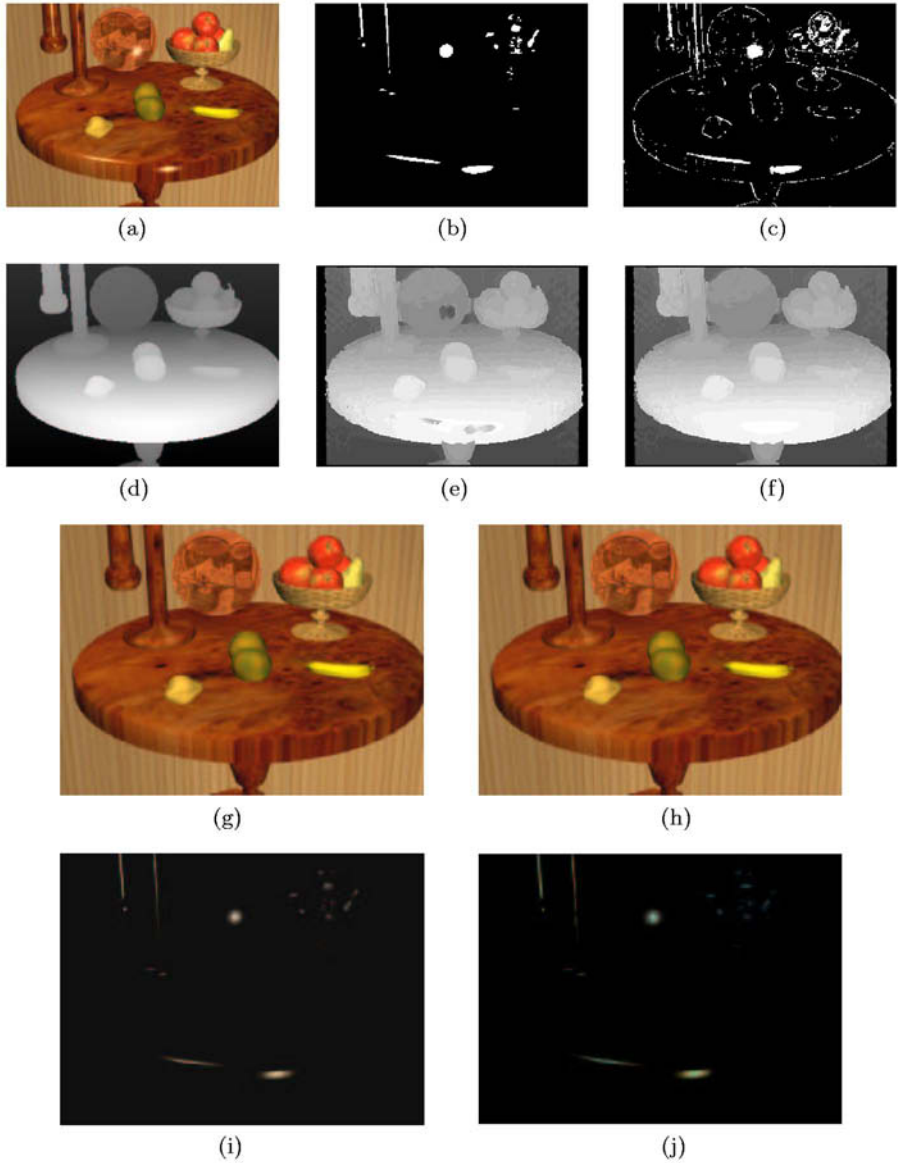
Fig. 4. Experimental results for synthetic scene: (a) original image; (b) ground truth detection of specular reflection, with white points representing specular and black ones representing diffuse; (c) our detection (d) ground truth depth; (e) depth estimation without special processing for specularities; (f) depth estimation using our approach; (g) ground truth diffuse component; (h) our separated diffuse component; (i) ground truth specular component; (j) our separated specular component
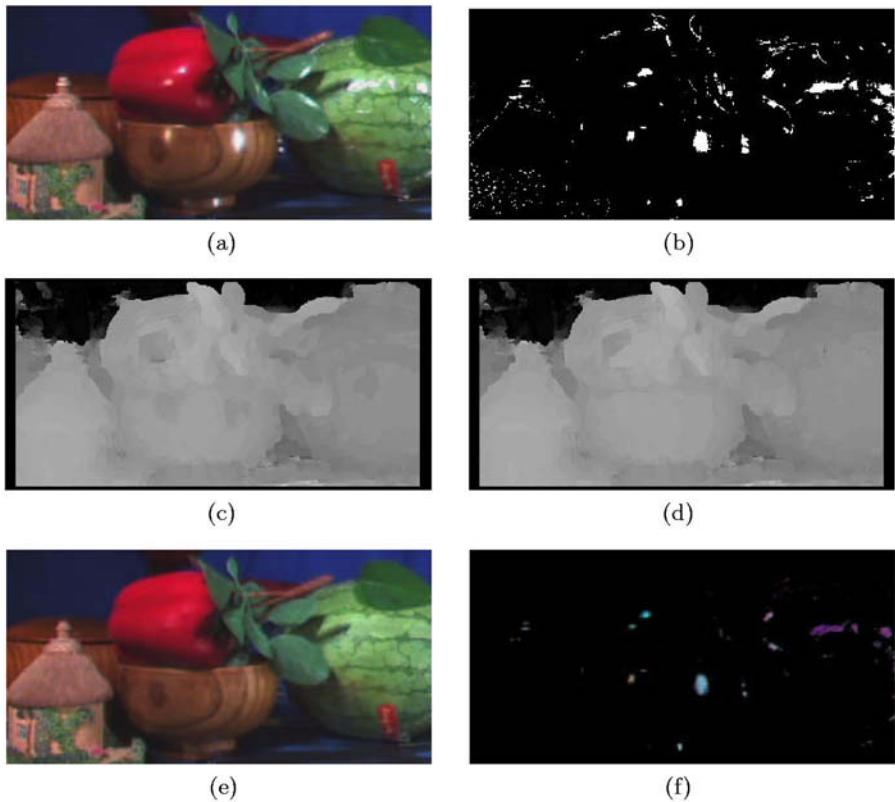
Fig. 5. Experimental results for real scene: (a) original image; (b) detection of specular reflection, with white points representing specular and black ones representing diffuse; (c) depth estimation without special processing for specularities; (d) depth estimation using our approach; (e) separated diffuse component; (f) separated specular component
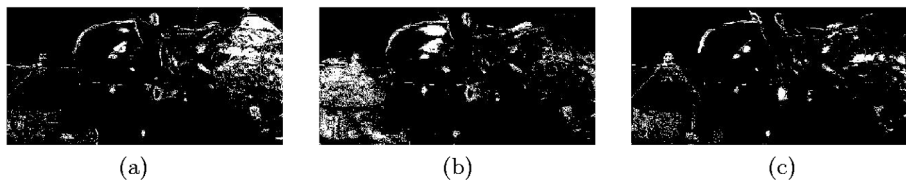

Fig. 6. Comparison of tri-view CHD with standard CHD (a) standard CHD between reference image and 10th image in sequence to the left; (b) standard CHD between reference image and 10th image in sequence to the right; (c) tri-view CHD

5 Discussion

In previous separation works, there is a strong dependence on segmentation to provide constraints on the diffuse component of specular reflections. Rather than relying on segmentation to get this information, we transfer this burden to stereo, for which this problem is more tractable. Separation methods require segmentation to group pixels into uniform-reflectance regions that contain both specular and diffuse reflections. Because of the large and often sharp intensity differences between the two forms of reflection, such a segmentation is difficult to compute. In stereo, the problem becomes one of matching equal-intensity diffuse pixels under the constraint of the specularity position, which should be easier than matching diffuse and specular pixels.

In the experimental results, color mixing was mentioned as an obstacle to detection, but was nevertheless processed correctly for the separation result. Other image effects that can disrupt our algorithm include color saturation, which can lead to crowding in a small part of the color histogram. Saturation can be substantially reduced or eliminated by capturing high dynamic range images [4]. Another obstacle is image noise, which can reduce accuracy in both detection and correspondence, but this can be remedied by averaging multiple images for each view.

Problems can also arise when our viewing assumptions are not held. Diffuse reflection from very rough surfaces may not follow the Lambertian model, and this is a hindrance for specularity detection and stereo correspondence in general. Our second and third assumptions, that specular color changes among viewpoints and that a specular scene point in one view is diffuse in some other views, are ordinarily broken when specular reflections exhibit little displacement among the stereo views. This may occur in areas of very high curvature and on surfaces with large roughness, because single pixels contain a broad range of surface normals in these cases. Although separation is difficult in such areas, the effect on stereo correspondence might not be substantial, since specularities behave similarly to fixed scene features.

6 Conclusion

We have described a new approach for identifying and separating specular components from an input image sequence. It integrates color analysis and multi-baseline stereo in a single framework to produce accurate separation and dense true depth. The color analysis, which identifies specular regions, is in the form of color histogram differencing with three new enhancements for increased robustness: (1) Extension to three views, (2) Use of epipolar constraints, and (3) Use of multiple triplets in a voting scheme. Once the specular pixels have been identified, we apply a dense stereo algorithm on the remaining pixels in order to extract the true depth. This stereo algorithm uses both shiftable windows and temporal selection in order to avoid the occlusion problem associated with depth discontinuity.

Currently the image sequences were acquired by moving the camera along a linear path with constant velocity. We plan to extend this work by using the

algorithm on sequences acquired with arbitrary camera motions. In addition, it would be interesting to learn how scene shape and lighting conditions affect the optimal camera motion for capture and subsequent specular extraction.

References

1. D.N. Bhat and S.K. Nayar. Binocular stereo in the presence of specular reflection. In *ARPA*, pages II:1305–1315, 1994.
2. D.N. Bhat and S.K. Nayar. Stereo in the presence of specular reflection. In *ICCV*, pages 1086–1092, 1995.
3. A.F. Bobick and S.S. Intille. Large occlusion stereo. *IJCV*, 33(3):1–20, Sept. 1999.
4. P.E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *Computer Graphics (SIGGRAPH)*, 31:369–378, 1997.
5. H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. Technical Report TR01-0017, UCLA, 2001.
6. S.B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In *CVPR*, pages 103–110, Dec. 2001.
7. G.J. Klinker, S.A. Shafer, and T. Kanade. A physical approach to color image understanding. *IJCV*, 4(1):7–38, Jan. 1990.
8. L. Lam and C.Y. Suen. Application of majority voting to pattern recognition: An analysis of its behaviour and performance. *IEEE Trans. on Systems, Man, and Cyberbetics*, 27(5):553–568, 1997.
9. S.W. Lee and R. Bajcsy. Detection of specularity using color and multiple views. *Image and Vision Computing*, 10:643–653, 1992.
10. D.C. Marr and T. Poggio. A computational theory of human stereo vision. In Lucia M. Vaina, editor, *From the Retina to the Neocortex: Selected Papers of David Marr*, pages 263–290. Birkhäuser, Boston, MA, 1991.
11. Y. Nakamura, T. Matsura, K. Satoh, and Y. Ohta. Occlusion detectable stereo - occlusion patterns in camera matrix. In *CVPR*, pages 371–378, 1996.
12. S.K. Nayar, X. Fang, and T.E. Boult. Removal of specularities using color and polarization. In *CVPR*, pages 583–590, 1993.
13. M. Okutomi and T. Kanade. A multiple baseline stereo. *IEEE PAMI*, 15, 1993.
14. Y. Sato and K. Ikeuchi. Temporal-color space analysis of reflection. *J. of the Opt. Soc. of America A*, 11, 1994.
15. S. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10, 1985.
16. F. Tong and B. V. Funt. Specularity removal for shape from shading. In *Proc. Vision Interface*, pages 98–103, 1988.
17. L.B. Wolff and T.E. Boult. Constraining object features using a polarization rflectance model. *IEEE PAMI*, 13, 1991.