Multibaseline Stereo in the Presence of Specular Reflections

Yuanzhen Li *
Chinese Academy of Sciences

Stephen Lin
Microsoft Research, Asia

Hanqing Lu
Chinese Academy of Sciences

Sing Bing Kang
Microsoft Research

Heung-Yeung Shum
Microsoft Research, Asia

Abstract

We address the problem of accurate depth estimation using multibaseline stereo in the presence of specular reflections. Specular reflections can cause the intensity and color of corresponding points to change dramatically according to different viewpoints, thus producing severe matching errors for various stereo algorithms. In this paper, we propose a new method to deal with this problem by treating specular reflections as occlusions. Our idea is to first detect specular pixels by computing the uncertainty of depth estimates. Then we combine the use of flexible windows and an adaptively selected subset of images to avoid these specular areas in all the multibaseline stereo images. Even though specularities may exist in the reference image, accurate depth is nevertheless estimated for all pixels. Experiments show that our consideration of specular reflections leads to improved stereo results.

1. Introduction

Stereo in the presence of specular reflection has been a challenging problem. Correspondence of points among stereo images relies heavily on the assumption of Lambertian reflectance, where each scene point has the same color or intensity in different views. Specular highlights in stereo images differ substantially from Lambertian reflection in that they change in position and color from view to view. This difference in behavior is disruptive to stereo matching, so we aim to circumvent these specularity effects.

A small amount of previous work addressed the problem of specular reflections in stereo. Bhat and Nayar [1] consider the likelihood of correct matching by analyzing relationship between stereo vergence and surface roughness. In [2] they further propose a trinocular system where two images are used at a time in computation of depth at a point. Jin *et al.* [5] poses this problem within a variational framework and seeks to estimate a 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.

In this paper, we propose a stereo approach that treats detected specular image regions as occlusions, since highlights effectively hide underlying Lambertian reflection. While methods have been presented to deal with occlusion due to scene geometry [7], they do not address the visibility problem posed by specular reflections. Geometric occlusions can occur when the view changes, but specular reflections typically appear in reference image itself. Moreover, shape and positioning of specular reflections in image sequences can differ greatly from those for geometric occlusions. To reduce the degradation in area-based correlation caused by specular highlights, we first detect them in stereo images using a correspondence uncertainty measure, and then disregard these specular points in computation of matching costs. We implicitly impose a depth continuity constraint between the specular pixel and its neighboring diffuse pixels by forming adaptive windows that ensure a certain number of diffuse pixels are included. Since specular points are rejected, matching windows are adaptive in size and have flexible shapes. They are used in a shiftable windows strategy which is effective in dealing with pixels near object boundaries and specular/diffuse boundaries. We also employ temporal selection, which has been proven effective in handling semi-occluded regions [7]. Because of 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.

We show in our experiments that the use of adaptive and shiftable windows, combined with temporal selection, greatly improves the matching of pixels that are specular in the reference image, pixels that are non-specular in the reference image but specular in others, and pixels that are near the boundary between specular and diffuse regions.

*This work was performed while the first author was visiting Microsoft Research, Asia.

2. Specularity detection by depth uncertainty

By forming specular masks for each image in a multi-baseline sequence, we can selectively aggregate support for depth estimation only from those pixels with high-confidence color, i.e. the diffuse pixels. Various detection algorithms have been developed that are based on physical models. They utilize physics-based cues such as structured light, color and polarization [4, 13, 10, 14, 8, 9]. For multi-baseline stereo, we propose an effective method of specularity detection based on uncertainty of depth estimates.

For a forward-facing multibaseline stereo configuration [11], 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

$$E_{SSSD}(x,y,d) = \sum_{k\neq 0} \sum_{(u,v)\in W(x,y)} \rho\big(I_0(u,v) - \hat{I}_k(u,v,d)\big), \tag{1}$$

where $\rho(\bullet)$ is 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 pixel (x, y) in reference image, the minimum E_{SSSD} for different disparity values determines the estimated disparity d and the uncertainty u of the estimation:

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

The value of uncertainty u is high when match quality is poor, as for a specular pixel. We use this quantity as a signal for specular reflection when its value exceeds a threshold t, set to the mean value of $u(x, y)$ plus one standard deviation. With this, the specular pixels can be represented as a binary image S determined by u:

$$S(x,y) = \begin{cases} 0 & \text{if } u(x,y) \leq t \\ 1 & \text{if } u(x,y) > t. \end{cases} \tag{2}$$

3. Adaptive and shiftable windows

With the detected specular pixels, we can perform more accurate stereo correspondence that excludes them from processing by using adaptive windows. Adaptive windows, proposed by Kanade and Okutomi [6], are square windows that extend by different amounts in each of four directions, to make sure that the window size is large enough to include enough intensity variation, but small enough to avoid the effects of projective distortion. We extend this idea of adaptive windows to exclude pixels detected as specular, and to also be sure it contains enough diffuse points for reliable matching. This results in windows not only adaptive in size, but also adaptive in shape.

In the formation of these adaptive windows, we use the specular detection S_k from (2) for respective images I_k. For each pixel (x, y), the window size n is first set to an initial value n_i, then n is extended until

$$C(x,y) = \sum_{(u,v)\in W_{n\times n}(x,y)} [1 - S_k(u,v)] \geq n^2 \times \alpha. \tag{3}$$

where α is a ratio which we set to 0.5. For correspondence between image pairs I_0 and I_k, we modify the $n \times n$ shiftable window $W_{n\times n}(x, y)$ into the window $W_f(x, y)$ whose support is flexibly shaped to exclude specularity:

$$W_f(x,y) \ = \ \big\{ \quad (u,v) \mid \ (u,v) \in W_{n\times n}(x,y) \ \ \& \\ S_0(u,v) = 0 \ \ \& \ \ S_k(u,v) = 0 \big\}.$$

We also implement this window as shiftable, using a separable sliding min-filter [12]. 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 handle geometric occlusions [3]. We show that integrated use of adaptive and shiftable windows 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 boundaries between specular and diffuse regions.

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

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

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

4. Temporal selection

We further extend our windowing procedure by dynamically selecting a subset of views where the support window is believed to be mostly diffuse and unoccluded. From the specular detection results, we can formulate a temporally selective aggregated matching error:

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

where

$$C(x,y) = \sum_{(u,v)\in W_f(x,y)} [1 - S_k(u,v)]. \tag{4}$$

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 for $E_{SSD}(x, y, d, k)$ 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 (4), and $wt(k) = 0$ for the remaining 50%. This temporal selection rule is similar to that described in [7].

Finally, we utilize a winner-take-all strategy to compute the final disparity:

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

5. Experimental Results

In this section, we present results on synthetic and real sequences to validate our approach.

For our experiments on synthetic images, we use eleven 320×240 images of a 57-image sequence generated using Phong shading. The baseline distance between consecutive views is 3.125mm. Our results are shown in Figure 1. (a) shows the reference image taken from our sequence. (b) shows the ground truth depth. (c) shows the specular mask we get from depth uncertainty, with white points representing specular and black ones representing diffuse. For comparison, we present the ground truth specular mask in (d). As can be seen, very few specular pixels are missed. There are some diffuse pixels falsely labelled as specular. One major reason is geometric occlusion. Another reason is color blending from more than one scene color imaged within a single pixel. However, the false labellings due to occlusions and color blending do not compromise the accuracy of our depth estimation, because these 'pseudo' specular pixels themselves are often origins of mismatches. By discarding them in the matching stage, we preserve the efficacy of our algorithm in handling specular reflections. Figures 1.(e-h) show the comparison of results of depth estimation using SSSD over fixed square windows with and without temporal selection, shiftable windows and our approach.

We present experimental results on two real sequences, shown in Figure 2 and Figure 3. Sequence A consists of 11 248×184 images and sequence B consists of 11 432×204 images , taken at regular intervals with a camera mounted on a horizontal translation stage, with the camera pointing perpendicularly to the direction of motion. In Figure 2 and Figure 3, (a) shows the reference image. (b) shows the specular mask we get using uncertainty of initial depth estimation. (c-d) show the comparison of stereo results using SSSD over fixed windows and our approach.

As exhibited in our experimental results, our stereo algorithm is effective in handling the problematic cases of specular pixels and pixels near specular/diffuse boundaries.

6. Discussion and conclusion

We presented an approach for reliable stereo in the presence of specular reflections by avoiding their detrimental effects. Stereo matching requires reflection to be Lambertian, and we treat specularities as occlusions of this diffuse reflection. Specular reflections are first detected by high uncertainty in their depth estimation. Then to handle their presence in a reference image, we perform among other views a diffuse point correspondence that is constrained by their disparity relationship to the highlight pixel. To account for specular reflections and occlusions among the stereo views, we presented extensions to adaptive and shiftable windows with temporal selection. These ideas were verified by experiments on synthetic and real scenes, which clearly exhibit the benefit of specularity processing.

A potentially attractive extension to our work would be to explicitly label specular pixels within a global energy minimization framework, and to reason about reflection state within this framework so that only truly diffuse pixels are matched.

References

[1] D. Bhat and S. Nayar. Binocular stereo in the presence of specular reflection. In *ARPA*, pages II:1305–1315, 1994.

[2] D. Bhat and S. Nayar. Stereo in the presence of specular reflection. In *ICCV*, pages 1086–1092, 1995.

[3] A. Bobick and S. Intille. Large occlusion stereo. *Int. J. of Computer Vision*, 33(3):1–20, Sept. 1999.

[4] G. Brelstaff and A. Blake. Detecting specular reflection using lambertian constraints. In *ICCV*, pages 297–302, 1988.

[5] H. Jin, A. Yezzi, and S. Soatto. Variational multiframe stereo in the presence of specular reflections. TR01-0017, UCLA, 2001.

[6] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. *PAMI*, 16(9):920–932, Sept. 1994.

[7] S. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. Technical Report MSR-TR-2001-80, Microsoft Research, Redmond, Sept. 2001.

[8] S. Lee and R. Bajcsy. Detection of specularity using color and multiple views. *Image and Vision Computing*, 10:643–653, 1992.

[9] S. Lin and S. W. Lee. A representation of specular appearance. In *In ICCV'99*, Sep. 1999.

[10] S. Nayar, X. Fang, and T. Boult. Removal of specularities using color and polarization. In *CVPR*, 1993.

[11] M. Okutomi and T. Kanade. A multiple baseline stereo. *IEEE TPAMI*, 15, 1993.

[12] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Technical Report MSR-TR-2001-81, MSR, Nov. 2001.

[13] S. Shafer. Using color to separate reflection components. *Color Research and Applications*, 10, 1985.

[14] L. Wolff and T. Boult. Constraining object features using a polarization rflectance model. *IEEE TPAMI*, 13, 1991.

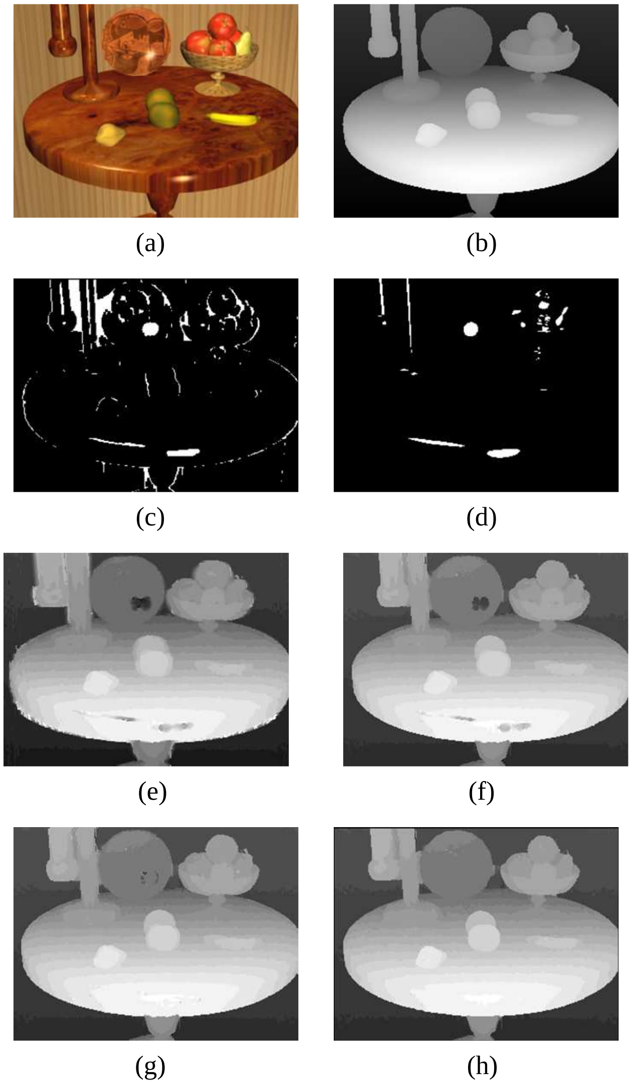

Figure 1. Experimental results for synthetic scene: (a) original image; (b) ground truth depth; (c) specular mask by depth uncertainty; (d) ground truth specular mask; (e-h) depth estimation results: (e) 3×3 centered square windows; (f) 3×3 centered square windows, with temporal selection; (g) 3×3 adaptive but non-shiftable windows, with temporal selection; (h) adaptive and shiftable windows, with temporal selection.

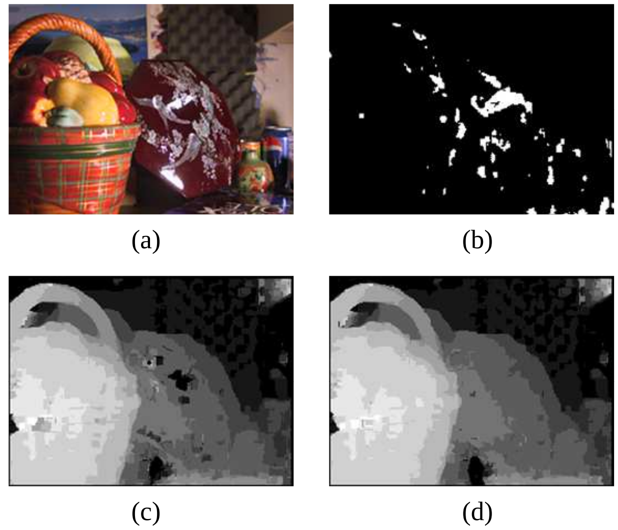

Figure 2. Experimental results for real scene A: (a) original image; (b) specular mask by depth uncertainty; (c-d) depth estimation results: (c) 5×5 centered square windows; (d) adaptive and shiftable windows, with temporal selection.

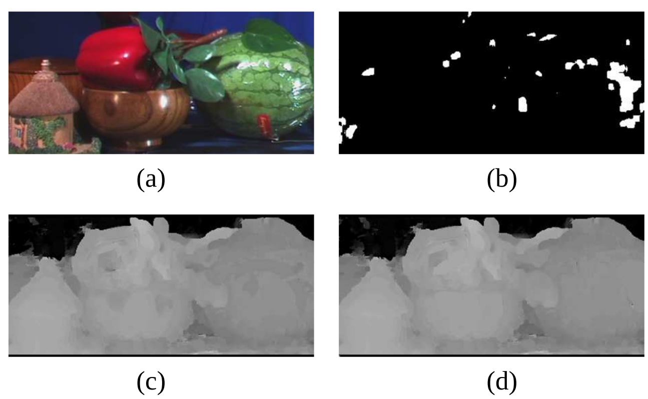

Figure 3. Experimental results for real scene B: (a) original image; (b) specular mask by depth uncertainty; (c) depth estimation results: (c) 7×7 centered square windows; (d) adaptive and shiftable windows, with temporal selection.