IMPROVED SPECULAR REGIONS LOCALIZATION AND OPTICAL-FLOW BASED MOTION ESTIMATION VIA JOINT PROCESSING

Ahmed S. Elliethy, Gaurav Sharma

Electrical and Computer Engineering Dept., University of Rochester, NY 14627, USA
{ahmed.s.elliethy, gaurav.sharma}@rochester.edu

ABSTRACT

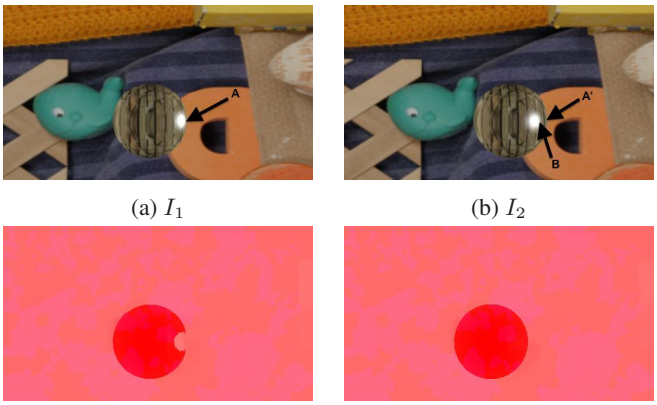
For specular regions (SRs), the assumption of brightness (or other) constancy between images corresponding to multiple views of a scene breaks down. As a consequence, optical-flow (OF) based motion-estimation (ME) algorithms that rely on constancy assumptions fail for specular regions. At the same time estimation of SRs in an image is also prone to errors, particularly to false positives from bright regions in the scene. In this paper, motivated by the fact that specular regions are typically encountered in image regions corresponding to portions of relatively smooth 3D surfaces, we propose an algorithm for improving ME and SRs localization via joint processing. Initial estimates of OF and of the SRs are obtained by conventional methods. The estimate of the SRs is updated using inconsistency of the OF with respect to the neighboring region to reinforce true positives and to reject false positives. The OF is then re-computed with a modified energy functional that, in effect, emphasizes regularization in a spatially adaptive neighborhood of the SRs to improve the estimated OF. Experimental results on synthetic and real image pairs demonstrate that the proposed algorithm offers a significant improvement in both SRs localization and ME over recently proposed methods for tackling these problems.

Index Terms— Specular region estimation, optical flow, motion estimation.

1. INTRODUCTION

An over-whelming majority of algorithms for computer vision (CV) are formulated under the assumption of Lambertian surfaces that contribute only diffuse reflection. Therefore, specular reflections, which are not uncommon in captured imagery, present a challenge for many computer vision tasks, such as image matching, color segmentation, and object recognition. Typically, the regions within images that correspond to specular reflections, viz. specular regions (SRs), are treated as outliers that are identified in a pre-processing stage and then excluded from downstream analysis by CV algorithms. However, methods for identifying the SRs within an image are also prone to significant false-positive (FP) errors from bright regions in the image. In this paper, we consider an alternative methodological approach in which we integrate the SRs estimation within the underlying CV algorithm. Specifically, we focus on OF based ME and demonstrate that by jointly computing estimates of OF and SRs, both estimates can be improved by leveraging information that each offers for the other.

Figure 1 demonstrates an image pair with SRs and the effect of the SRs on ME. Based on brightness or texture constancy assumption, OF estimation algorithms match the SR pixels in one image with SR pixels in the other image regardless of the true motion for these pixels. For example, the pixel labeled A in the left image is matched incorrectly with the pixel labeled B in the right image instead of the correct match A'. Figure 1(c) illustrates the map of MEs (in standard color encoded format) obtained for this pair with a state-of-the-art OF estimation method [1], where the error in the SR is clearly apparent. Figure 1(d) illustrates the result of our proposed method where the accuracy is very significantly improved in the SRs.



(a) I_1 (b) I_2

(c) Optical Flow [1] (d) Optical Flow (proposed)

Fig. 1: Effect of specular reflection on optical flow estimation on Ball image pair: (a), (b) are the input images, (c) is a color coded optical flow obtained by [1], while (d) is our result.

A number of methods have been previously proposed for the detection and removal of SRs. A number of these require specialized equipment or setup. For instance, different polarization angles [2], or different light sources (flash) positions [3, 4] have been used to distinguish between the SRs and diffuse reflecting regions. Under a constant linear motion constraint, the epipolar constraint has also been used in [5] to identify SRs. In [6], the illuminant chromaticity, correspondence, and SRs are estimated in a single framework, which, however, encounters a challenge with saturated pixels commonly encountered in SRs. In [7], an iterative method was proposed for both estimate the illuminant chromaticity and enhance the OF. The method estimate the illuminant chromaticity from the regions where OF fails, then projecting the input images to a color space that is perpendicular to the estimated illuminant chromaticity to enhance the OF estimation. However, this color projection will fail for achromatic surfaces and saturated pixels, and also as shown in Fig. 1, the specular pixel may be assigned to another specular one in the other view, and that will result in a small difference between them

This work was supported in part by the National Science Foundation grant CNS-1239423.

which will not be detectable for the illuminant chromaticity estimation process.

Our key observation in this paper is that, SRs in images typically occur in image regions corresponding, in the 3D scene, to relatively smooth surfaces and therefore, the true motion corresponding to the SR pixels is typically consistent with the neighboring region within the same object whereas the estimated motion is typically inconsistent with the neighboring region. This inconsistency can therefore be used to vet estimates of SRs and eliminate false positives improving the estimates of the SRs. In turn, the improved estimates of SRs, can improve OF based ME estimates by downplaying the data term that causes spurious matches in this region. Thus, we obtain better estimates of both the OF and of the SRs via a joint processing strategy that we detail in the following sections.

This paper is organized as follows. Section 2 outlines the standard OF estimation formulation and highlights how SRs pose a challenge. Section 3 explains our proposed algorithm that treats the OF estimation and SRs localization problems jointly with each benefiting from the other. Results and a comparison against alternative methods are presented in Section 4. We conclude the paper in Section 5.

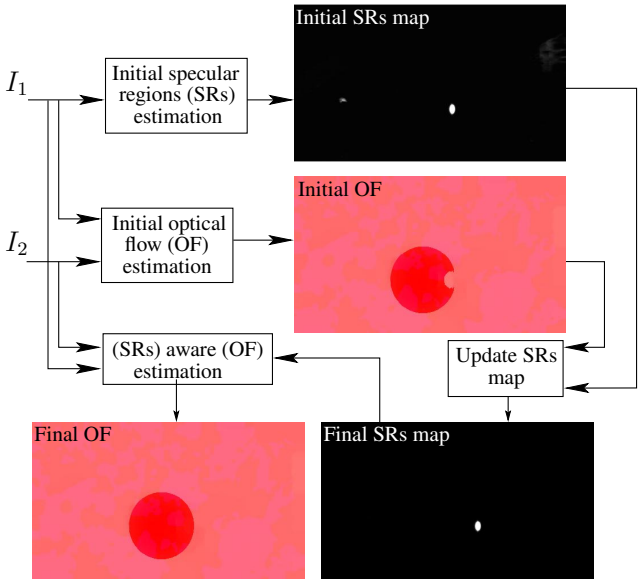


Fig. 2: Proposed algorithm for joint estimation of OF and SRs.

2. SPECULAR REGIONS INDUCED CHALLENGES FOR OF ESTIMATION

Given an image pair[1] (I_1, I_2), the OF estimation can be posed as an energy minimization problem as

$$\mathbf{d}^*(\mathbf{x}) = \arg\min_{\mathbf{d}(\mathbf{x})} E(\mathbf{d}(\mathbf{x})), \tag{1}$$

where $\mathbf{d}(\mathbf{x})$ is the disparity field defined on sampling grid of I_1. The objective function can be written in a discrete form as

$$\begin{aligned} E(\mathbf{d}(\mathbf{x})) &= E_D(\mathbf{d}(\mathbf{x})) + \lambda E_S(\mathbf{d}(\mathbf{x})) \\ &= \sum_{\mathbf{x}} \Phi_D(F_1(\mathbf{x}) - F_2(\mathbf{x}+\mathbf{d}(\mathbf{x}))) + \lambda \sum_{\mathbf{x}} \Phi_S(\mathbf{d}(\mathbf{x})), \end{aligned} \tag{2}$$

where Φ_D represent the *data penalty function* that enforces the constancy assumption, Φ_S is the *smoothness penalty function* that enforces the spatial regularity (coherence) of the estimated OF, and F_1 & F_2 are spatial feature representations corresponding to I_1 & I_2, respectively. In their simplest form, F_1 & F_2 could correspond directly to the image intensities for I_1 & I_2; alternatively, they may be computed as dense feature descriptors, for example SIFT feature descriptors [8]. The objective function (2) was first proposed with F_1 & F_2 as image intensities in [9]. Recently, this formulation gives very competitive results by using different penalty functions and different feature descriptors for F_1 & F_2 [1]. However, using the formulation of (2) in the presence of SRs yields spurious motion estimate for these regions as Φ_D is typically minimized by matching SRs pixels in the two images as shown in Fig. 1. The problem can be partly mitigated by estimating the SRs, however, SRs estimation methods are prone to errors, in particular to FPs in bright regions of the image. To address, this challenge we propose a joint algorithm to enhance the OF estimation and, simultaneously, obtain a better estimate for the SRs.

3. PROPOSED ALGORITHM

A high level overview of the proposed algorithm is shown in block-diagram format in Fig. 2 using an illustrative example with the images I_1 & I_2 shown in Fig. 1. The proposed method is based on two ideas. First, because SRs typically occur over smooth surfaces in 3D space, interpolation of the motion field for the neighboring regions *within the same object* can provide a good estimate of the motion field for SRs. Secondly, the spurious motion estimates for SRs that arise due to the data term in (2) are typically inconsistent with the motion estimates obtained by the interpolation process whereas for bright non-specular regions, this is typically not the case. We use the inconsistency of the motion field with the motion in surrounding regions within the same object as a positive reinforcement for identifying SRs that allows FPs to be eliminated.

Specifically, we propose a modified energy function

$$\begin{aligned} E(\mathbf{d}) = &\sum_{\mathbf{x}} \gamma_S(\mathbf{x}) \left(\Phi_D(F_1(\mathbf{x}) - F_2(\mathbf{x}+\mathbf{d}(\mathbf{x})))\right) + \\ &\lambda \sum_{\mathbf{x}} \Phi_S(\mathbf{d}(\mathbf{x})), \end{aligned} \tag{3}$$

where $\gamma_S(\mathbf{x})$ represents a spatially adaptive weighting for the data term. The impact of the error introduced in the ME by the data term in SRs can therefore be eliminated if this weighting takes a value of 1 over diffuse regions and 0 in the SRs with a smooth transition in intermediate regions. The data weighting factor $\gamma_S(\mathbf{x})$ results in only a minor change to the corresponding Euler-Lagrange equations that characterize the minimum for (3), and also do not depend on the specific choice of the Φ_D or Φ_S penalty functions.

For the energy function in (3) to be effective, the weighting $\gamma_S(\mathbf{x})$ needs to be determined using the locations for SRs in I_1, i.e. the specularity map. We obtain an initial coarse estimate of the specularity map using a dark-channel prior as in [10] and improve this by eliminating FPs by leveraging information from OF

[1] Intensity values in the images I_1, and I_2 are assumed to be normalized to a $[0, 1]$ range.

estimation. Based on the observation that diffuse pixels are characterized by a low intensity level in at least one color channel among RGB channels, we can use the dark channel, defined by, $I_d(\mathbf{x}) = \min_{c\in\{r,g,b\}} I_1^c(\mathbf{x})$, as a coarse estimate for the specular reflections of an image [10]. Specifically, a coarse SRs binary map, identifying the regions with significant specularity, is obtained by thresholding

$$\gamma_S^0(\mathbf{x}) = 1 - e^{\frac{I_d(\mathbf{x})-1}{\sigma_s}}. \tag{4}$$

To address the FPs typical with this coarse estimate, the SRs binary map is segmented into a list of connected components (CCs) and the specularity within each CC validated by testing the consistency of the motion within the component with that of a small neighborhood that is adaptively determined to lie within the same object as the CC in question. It is desirable to have this adaptive region be confined to the same object over which the putative SR CC lies. To accomplish this, inspired by active contours [11], we define a temporally evolving level set function $\varphi(\mathbf{x}, t)$ where t indicates the time (iteration number) variable. The function $\varphi(\mathbf{x}, 0)$ is initialized to values of 1 and -1, respectively, inside and outside of the putative SR CC. The enlarged SR region for the CC at time t is then represented by the zero level set $\varphi(\mathbf{x}, t) = 0$, where the standard level set equation [11]

$$\frac{\partial\varphi}{\partial t} = V(\mathbf{x})\,|\nabla\varphi| \tag{5}$$

is used to evolve the level set function with the velocity term $V(\mathbf{x})$ defined as

$$V(\mathbf{x}) = \frac{1}{\alpha\|\nabla\sigma(\mathbf{x})\|^2 + 1} + \frac{1}{\beta\|\Delta\sigma(\mathbf{x})\| + 1}, \tag{6}$$

where

$$\sigma(\mathbf{x}) = \frac{I_1(\mathbf{x})}{I_1^r(\mathbf{x}) + I_1^g(\mathbf{x}) + I_1^b(\mathbf{x})} \tag{7}$$

is the *normalized rgb* or *chromaticity* of the image $I_1(\mathbf{x})$ [12], and α, β are positive constants. The evolution of the level function correspondingly enlarges the contour whose inside corresponds to the CC along the normal direction $\vec{N}(\mathbf{x})$ to the contour (at each point $\mathbf{x}$) with the velocity $V(\mathbf{x})$. The velocity in (6) is motivated by the fact that adjoining SRs, the chromaticity varies smoothly within the same object whereas the change is abrupt across object boundaries. The form of $V(\mathbf{x})$ is designed to stop expansion of the boundary when an edge is encountered and to allow expansion in other directions that are constant or smoothly varying. The region between the original boundary of the CC and the adaptively expanded boundary is denoted by B_i. The CC is then validated as a true SR if the difference between the averages of the MEs in the CC and in the boundary region B_i are smaller than an empirically determined threshold τ_2, otherwise the CC is considered a FP and dropped from the estimate of the SRs.

The refined estimate of the SRs obtained by the afore-mentioned procedure is denoted by a binary indicator function $\lambda_s(\mathbf{x}) \in \{0, 1\}$ which is incorporated in our adaptive weighting function $\gamma_S(\mathbf{x})$ by first computing

$$\tilde{\gamma}_S(\mathbf{x}) = 1 - \lambda_s(\mathbf{x})e^{\frac{I_d(\mathbf{x})-1}{\sigma_s}}. \tag{8}$$

Then by evolving $\tilde{\gamma}_S(\mathbf{x})$ using the velocity in (6), we obtain the final weighting $\gamma_S(\mathbf{x})$ that is implicitly gives more trust to the data term as we go far from the SR. Algorithm 1 summarizes the entire algorithm. The parameter τ_1 is determined automatically using the method in [13], and we estimate τ_2 empirically.

Algorithm 1: Proposed algorithm for jointly estimating optical flow and specularity map

Data: Input F_1, F_2
Result: $\mathbf{d}(\mathbf{x}), \lambda_s(\mathbf{x})$
Optimize (3) with $\gamma_S(\mathbf{x}) = 1$ to get initial ME $\mathbf{d}^{\mathrm{o}}(\mathbf{x})$;
Compute $\gamma_S^{\mathrm{o}}(\mathbf{x})$ as in (4);
Identify potential SRs $\mathcal{R} = \{\mathbf{x} : \gamma_S^{\mathrm{o}}(\mathbf{x}) < \tau_1\}$;
Segment $\mathcal{R}$ into a list of connected components $\mathcal{G}$;
for *Each connected component* $\mathcal{C}_i$ *in* $\mathcal{G}$ **do**
    $A_i = avg(\mathbf{d}^{\mathrm{o}}(\mathbf{x}))$ for $\mathbf{x} \in \mathcal{C}_i$;
    Get a small region B_i around $\mathcal{C}_i$ by evolve $\mathcal{C}_i$'s boundary using $V(\mathbf{x})$ with velocity in (6);
    $A_i^b = avg(\mathbf{d}^{\mathrm{o}}(\mathbf{x}))$ for $\mathbf{x} \in B_i$;
    if $\|A_i - A_i^b\|_2 > \tau_2$ **then**
        add i to the specularity set $\mathcal{S}$;
    end
end
if $\mathcal{S} = \phi$ **then**
    return;
end
$\lambda_s(\mathbf{x}) = 1$, for $\mathbf{x} \in \mathcal{S}$ and 0 o.w.;
Compute $\tilde{\gamma}_S(\mathbf{x})$ as in (8);
Expand $\tilde{\gamma}_S(\mathbf{x})$ using $V(\mathbf{x})$ in (6) to obtain $\gamma_S(\mathbf{x})$;
Optimize (3) to get final ME $\mathbf{d}(\mathbf{x})$;

4. EXPERIMENTS

We demonstrate our results using two of the state of the art methods for OF estimation [1, 8] using both semi-synthetic data for which ground truth is available and allows numerical validation and for captured imagery with SRs. For the semi-synthetic data, we adopt an approach similar to [14]. We use the Middlebury data set (*www.middlebury.edu/stereo*) [15] with the available ground truth disparity to generate some test images with specularity introduced using Blender 3D modeling environment [16]. These image pairs are shown in Fig. 3 (a). This methodology allow us to use the ground truth disparity map to provide quantitative measure of performance, and additionally uses diffuse colors from actual scenes making the scenario close to reality.

Our results is shown in Fig. 3 (b), (c) using the standard color encoding [17] for the estimated motion vectors obtained from OF[2]. Also, Table 1 shows a comparison of both average end-point error (EPE) and angular error (AAE) of the original methods [1, 8] (**w/o SRs handling**) with the method in [7] and our algorithm. In order to make our accuracy measures unbiased with respect to the image size and the amount of specularities contained on the input images, we report both EPE, and AAE over a rectangular window around the specular reflecting regions that extends 20 pixels beyond pixels with specular reflection (from known ground truth). The method in [7] fails when the input images contain many achromatic surfaces because upon color projection along the direction orthogonal to the illuminant chromaticity little energy is retained. As a result, the methodology of [7] gives a higher EPE and AAE. This validate our approach of discounting the data term in OF estimation for the SRs. Also, Fig. 4 shows our results[3] on some indoor images that lit by the sun light enters through a window. It is clear that there is

[2]A value of $\tau_2 = 0.5$ was used in conjuction with the OF method [1] and $\tau_2 = 0.2$ was used with [8].

[3]The threshold τ_2 was set here by trial and error.

(a) Input : Cones and Sawtooth image pairs

(b) Results for Cones image pair using [8] for OF estimation

(c) Results for Sawtooth image pair using [1] for OF estimation

Fig. 3: Input image pair and our result. In (b)-(c) columns from left to right represent: (1) the initial specularity map, (2) the initial optical flow, (3) the final specularity map, (4) the final optical flow, (5) ground truth optical flow. Images are best viewed electronically.

	OF	w/o SRs handling		Method in [7]		Our algorithm	
		AAE	EPE	AAE	EPE	AAE	EPE
Ball	[1]	0.074	0.583	0.448	0.805	**0.029**	**0.120**
	[8]	0.106	1.799	**0.093**	1.621	0.095	**1.441**
Cones	[1]	2.568	0.758	4.788	1.238	**0.673**	**0.356**
	[8]	9.062	3.051	7.718	1.930	**0.444**	**0.587**
Sawtooth	[1]	5.217	0.649	6.104	0.631	**0.825**	**0.104**
	[8]	3.613	0.689	6.735	0.574	**0.490**	**0.148**

Table 1: Quantitative comparison between our method and the original one.

a noticeable enhancement for both the final specularity map and the estimated OF compared to initial one. Similar results were obtained for additional image pairs with SRs captured under different sources of illumination but are not included here due to space constraints.

5. CONCLUSION

We propose a method that jointly improves the estimation of optical flow (OF) and of specular regions (SRs) from a stereo image pair with specular reflections. By exploiting information from each of the two components for the other, we demonstrated that both estimates can be significantly improved. Our estimates of OF offer a significant improvement over both methods that disregard SRs and over a recent alternative method that considers the specular reflection challenges for OF.

Fig. 4: Results on actual captured images using [8] for OF estimation. Left column shows the input images, middle shows the initial and final OF, and right one shows initial and final specularity map

6. REFERENCES

[1] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2010, pp. 2432–2439.

[2] S. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *IEEE Intl. Conf. Comp. Vision.*, vol. 21, no. 3, pp. 163–186, 1997.

[3] R. Feris, R. Raskar, K.-H. Tan, and M. Turk, "Specular reflection reduction with multi-flash imaging," in *Proc. 17th Brazilian Symp. on Comp. Graphics and Image Proc.*, Oct 2004, pp. 316–321.

[4] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. on Graphics*, vol. 24, no. 3, pp. 828–835, Jul. 2005.

[5] S. Lin, Y. Li, S. B. Kang, X. Tong, and H.-Y. Shum, "Diffuse-specular separation and depth recovery from image sequences," in *Proc. European Conf. Computer Vision*, London, UK, UK, 2002, pp. 210–224.

[6] Q. Yang, S. Wang, N. Ahuja, and R. Yang, "A uniform framework for estimating illumination chromaticity, correspondence, and specular reflection," *IEEE Trans. Image Proc.*, vol. 20, no. 1, pp. 53–63, Jan 2011.

[7] C. Arora and M. Werman, "Optical flow for non Lambertian surfaces by cancelling illuminant chromaticity," in *IEEE Intl. Conf. Image Proc.*, Oct 2014, pp. 1977–1981.

[8] C. Liu, J. Yuen, and A. Torralba, "SIFT flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 33, no. 5, pp. 978–994, May 2011.

[9] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[10] H. Kim, H. Jin, S. Hadap, and I. Kweon, "Specular reflection separation using dark channel prior," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, June 2013, pp. 1460–1467.

[11] S. Osher and R. P. Fedkiw, *Level set methods and dynamic implicit surfaces*. Springer Verlag, 2003, vol. 153.

[12] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, N.J.: Prentice-Hall, 1989.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Sys., Man, and Cyber.*, vol. 9, no. 1, pp. 62–66, Jan 1979.

[14] W. Zhou and C. Kambhamettu, "Binocular stereo dense matching in the presence of specular reflections," in *IEEE Intl. Conf. Comp. Vision, and Pattern Recog.*, vol. 2, 2006, pp. 2363–2370.

[15] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Intl. J. Computer Vision*, vol. 47, no. 1-3, pp. 7–42, Apr. 2002.

[16] Blender Online Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Blender Institute, Amsterdam. [Online]. Available: http://www.blender.org

[17] S. Baker, S. Roth, D. Scharstein, M. Black, J. Lewis, and R. Szeliski, "A database and evaluation methodology for optical flow," in *IEEE Intl. Conf. Comp. Vision.*, Oct 2007, pp. 1–8.