THREE-VIEW DENSE DISPARITY ESTIMATION WITH OCCLUSION DETECTION

Xiaodong Huang and *Eric Dubois*

School of Information Technology and Engineering (SITE)
University of Ottawa
Ottawa, ON, K1N 6N5 Canada
Email: {xhuang, edubois}@site.uottawa.ca

ABSTRACT

An algorithm for disparity estimation with energy-based variational regularization using three image views is presented. In the new algorithm, dense disparity maps, including the disparity values for feature points, are first estimated by phase-based methods. Then a set of coupled partial differential equations (PDEs) are solved to refine these disparity maps together with the feature information. The visual appearance of the refined disparity is better than that of the results using similar variational regularization involving only one stereo pair of images. In addition, a new effective occlusion detection scheme is presented with good results.

1. INTRODUCTION

Recently variational regularization and partial differential equation (PDE) approaches have been increasingly applied to disparity estimation of stereo and multiview images [1][2]. In [1] as well as other similar approaches for small-baseline stereo, usually a coarse disparity map estimated by area-based methods like the block correlation is used as initialization to the PDE, which is solved iteratively as a refinement process for the disparity. The refinement can give good results under the condition that the coarse disparity has enough accuracy; otherwise the iterative process in solving the PDE is very likely to fall into local minima. However, the area-based methods for disparity estimation are not robust. They will cause problems in areas with strong features as well as in areas with slanted surfaces, and are also sensitive to brightness variations. Another inconvenience for the area-based methods is that the block size needs to be adjusted for images with different feature densities.

During the last decade, another class of approaches for disparity estimation called *phase-based methods* has been developed, e.g., [3]. The methods which belong to this class use the local phase information of the images to find the disparity. They are more robust to different texture levels and slanted surfaces, less sensitive to brightness variations, and could give better results than the area-based methods without adjusting the support region of each Gabor filter.

Most of the algorithms for disparity estimation only deal with the case of two-view stereo. As shown in [4][5], one approach to increase the precision of disparity estimation for small-baseline stereo is to use multiple-baseline or multiview stereo sequences. We use this approach with three equally spaced images in this paper. However, instead of using the area-based sum of sum of squared-difference (SSSD) as used in [4][5], we use the robust phase-based methods combined with some feature information to obtain a coarse disparity map. For the refinement, unlike the approaches in [4][5], we are using the variational PDE approach which takes into account the inter-relation of different disparity maps for the refinement of the disparity, and treats the feature information separately in the refinement process to keep the object boundaries clear. The difference of our approach with that of in [2] is that we are dealing with small-baseline parallel stereo sequences, while the approach in [2] handles wide-baseline multiview images with quite different directions and therefore does not take into account the inter-relation of different disparities as a constraint. This paper presents this novel disparity estimation algorithm based on three views along with a new occlusion detection scheme based on the results of the refined disparity.

2. PROBLEM FORMULATION

In this paper, we suppose that three images are taken along a straight line with approximately equal distance between any two consecutive camera positions. The optical axes are perpendicular to the straight line and thus result in an equi-baseline stereo sequence containing three images. As shown in Fig. 1, let the three images be denoted by I_l, I_{r1} and I_{r2}, from left to right. $d_1(x, y)$ represents the disparity map for the pixels in I_l to I_{r1}, i.e. $I_l(x, y) \cong I_{r1}(x - d_1(x, y), y)$, as indicated in Fig. 1 for $d_1(x, y)$ with the arrow; there is a similar definition for the other disparity maps $d_2(x, y)$ and $d_3(x, y)$. We assume that the vertical disparity is zero. In two-view stereo matching, to estimate $d_1(x, y)$, only one image pair (I_l and I_{r1}) is used. Now, with the additional image I_{r2}, we will try to increase the precision

0-7803-9134-9/05/$20.00 ©2005 IEEE

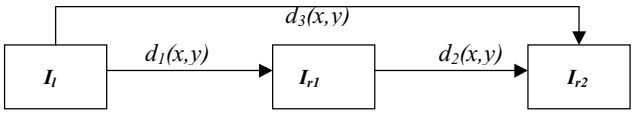


Fig. 1. Multi-base stereo with three images

of $d_1(x, y)$ by using the information from the other disparity maps. Because these disparity maps are inter-related by $d_1(x, y) + d_2(x - d_1(x, y), y) = d_3(x, y)$, this coupling relation can bring additional constraints to the refinement process and should increase the accuracy of the estimate of $d_1(x, y)$.

3. DISPARITY ESTIMATION WITH VARIATIONAL REFINEMENT

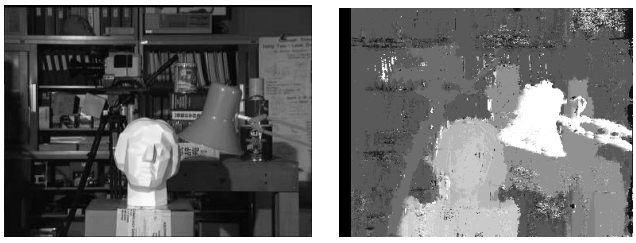
For the estimation of $d_1(x, y)$, $d_2(x, y)$ and $d_3(x, y)$, we first do the coarse matching for each of them separately using the Gabor transform. Then the refinement will be performed using a combined energy functional containing $d_1(x, y)$, $d_2(x, y)$ and $d_3(x, y)$.

3.1. Stereo Matching Using Gabor Transform

The method that we used for a coarse estimation of the disparity using the Gabor transform is mainly based on the algorithm in [3], in which a set of quadrature-pair Gabor filters are used. Each quadrature-pair Gabor filter is a set of discretized samples of a Gabor function with different tuning frequencies, and is used for the filtering of the stereo images to obtain the approximate Gabor transform coefficients at those frequencies. In addition to Gabor filtering, the images are also filtered by Sobel masks to obtain the Sobel coefficients of each pixel. Assume that the outputs of k^{th} Gabor filter pair are $G_l^k(x, y)$ and $G_r^k(x, y)$ for the left and right images respectively; the Sobel coefficients are $S_{l\_x}(x, y)$ and $S_{r\_x}(x, y)$ in x direction, $S_{l\_y}(x, y)$ and $S_{r\_y}(x, y)$ in y direction, for the left and right images respectively. Then the disparity $\hat{d} \in [0, d_{max}]$ for a position (x, y) in the left image is determined as:

$$\begin{aligned}\hat{d}(x,y) = \arg\min_d \Big\{ \sum_k \big[&|Re\{G_l^k(x,y)\} - Re\{G_r^k(x-d,y)\}| \\ &+ |Im\{G_l^k(x,y)\} - Im\{G_r^k(x-d,y)\}| \big] \\ &+ w_s \big[|S_{l\_x}(x,y) - S_{r\_x}(x-d,y)| \\ &+ |S_{l\_y}(x,y) - S_{r\_y}(x-d,y)| \big] \mathcal{I}_\tau(S_l) \Big\} \qquad (1)\end{aligned}$$

where $Re\{G_l^k(x, y)\}$ and $Im\{G_l^k(x, y)\}$ are the real and imaginary parts of $G_l^k(x, y)$, and similarly for $G_r^k(x, y)$; $\mathcal{I}_\tau(S_l)$ is the indicator function which indicates if $S_l = \sqrt{S_{l\_x}^2 + S_{l\_y}^2}$ is larger than a threshold τ, and w_s is a weighting coefficient. Therefore, for those edge pixels in the left image whose Sobel coefficients are larger than τ, their disparities are estimated by comparing Sobel coefficients in addition to Gabor coefficients. This made the disparities for such pixels more reliable, while the disparities for non-edge pixels are estimated only based on their Gabor coefficients. The reason for distinguishing feature points is that we wish to suppress the smoothing of these pixels in the refinement stage in order to keep the object contours clear.



(a) (b)

Fig. 2. (a) Original left image of *Tsukuba*. (b) The coarse disparity estimated by the phase method.

(a) (b)

Fig. 3. (a) Original left image of *Flower Garden*. (b) The coarse disparity estimated by the phase method.

We refer to [6] for the implementation details of the Gabor filters used in this paper. In our experiments, we used two image sequences – *Tsukuba* and *Flower Garden* – both of which are taken along a straight line and are approximately equi-distant for any two consecutive images. We choose frames 1 – 3 from *Tsukuba* and frames 21, 23 and 25 from *Flower Garden*, with the first frame as left image and the last frame as second right image for each sequence. The original left images of *Tsukuba* and *Flower Garden* are shown in Fig. 2(a) and Fig. 3(a) respectively. We used $w_s = 10$ and $\tau = 0.4$. The disparity fields $d_1(x, y)$, $d_2(x, y)$ and $d_3(x, y)$ are estimated separately using (1). The results of $d_1(x, y)$ are shown in Fig. 2(b) and Fig. 3(b) for *Tsukuba* and *Flower Garden*.

3.2. Refinement with Variational Regularization

The data fidelity term for $d_1(x, y)$ and $d_3(x, y)$ which is part of the energy functional we need to minimize in a multi-

baseline framework is

$$E_D(d_1,d_3;d_2)=\iint\big\{[I_l(x,y)-I_{r1}(x-d_1(x,y),y)]^2 + [I_l(x,y)-I_{r2}(x-d_3(x,y),y)]^2+[I_{r1}(x-d_1(x,y),y) - I_{r2}(x-d_1(x,y)-d_2(x-d_1(x,y),y),y)]^2\big\}\mathrm{d}x\,\mathrm{d}y \tag{2}$$

where the integration is with respect to the left image I_l. There are three squared terms in the above integration. The first term is just the data fidelity term used in disparity estimation involving only one stereo pair, while the following two extra data fidelity terms come from the additional third image for our multi-baseline stereo case. Since disparity estimation is an ill-posed inverse problem, additional constraints are needed to regularize the disparity estimates. We used variational regularization to control the smoothing of disparity variations.

Variational regularization has had extensive application in optical flow estimation, as well as for disparity estimation [1][2]. The idea is to control the smoothing of disparity variations using the values of image gradients. When the value of image gradient is low the disparity would be smoothed, while the smoothing process is stopped when the value of image gradient is high, which represents a possible object boundary. The functional we used for the regularization of $d_1(x,y)$ is

$$E_R(d_1)=\iint\Big\{\frac{1}{(1+I_{l,x}^2)^2}d_{1,x}^2+\frac{1}{(1+I_{l,y}^2)^2}d_{1,y}^2\Big\}\mathrm{d}x\,\mathrm{d}y \tag{3}$$

where $d_{1,x}$ and $d_{1,y}$ are derivatives of $d_1(x,y)$ in x and y directions respectively, and similarly for $I_{l,x}$ and $I_{l,y}$. Since the disparity $d_3(x,y)$ is also directly involved in (2) with respect to the left image I_l, another regularization term which is the same as (3) but with respect to $d_3(x,y)$ is also used. In addition, the inter-relation among $d_1(x,y)$, $d_2(x,y)$ and $d_3(x,y)$, i.e. $d_1(x,y)+d_2(x-d_1(x,y),y)=d_3(x,y)$, brings further constraint to the three disparity maps. Thus, we can write the final regularization functional for the data fidelity term (2) as

$$E_R(d_1,d_3;d_2)=E_R(d_1)+E_R(d_3) + \alpha\iint[d_1(x,y)+d_2(x-d_1(x,y),y)-d_3(x,y)]^2\mathrm{d}x\,\mathrm{d}y. \tag{4}$$

The overall energy functional to be minimized for $d_1\,(x,y)$ and $d_3(x,y)$ is

$$E(d_1,d_3;d_2)=E_D(d_1,d_3;d_2)+\lambda E_R(d_1,d_3;d_2), \tag{5}$$

where α and λ in (4) and (5) are two regularization parameters.

The minimization of $d_1(x,y)$ and $d_3(x,y)$ is achieved by applying a gradient descent method to solve the associated Euler-Lagrange equations of (5) with respect to $d_1(x,y)$ and $d_3(x,y)$:

$$\begin{aligned}\frac{\partial d_1}{\partial t}=&[I_l(x,y)-I_{r1}(x-d_1,y)]\times I_{r1,x}(x-d_1,y)\\ &+[I_{r1}(x-d_1,y)-I_{r2}(x-d_1-d_2(x-d_1,y),y)]\\ &\times\big\{I_{r2,x}(x-d_1-d_2(x-d_1,y),y)\\ &\quad\times[1-d_{2,x}(x-d_1,y)]-I_{r1,x}(x-d_1,y)\big\}\\ &+\lambda\alpha[d_1+d_2(x-d_1,y)-d_3]\times[1-d_{2,x}(x-d_1,y)]\\ &-\lambda\Big\{\frac{\partial}{\partial x}\Big[\frac{d_{1,x}}{(1+I_{l,x}^2)^2}\Big]+\frac{\partial}{\partial y}\Big[\frac{d_{1,y}}{(1+I_{l,y}^2)^2}\Big]\Big\},\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{\partial d_3}{\partial t}=&[I_l(x,y)-I_{r2}(x-d_3,y)]\times I_{r2,x}(x-d_3,y)\\ &-\lambda\alpha[d_1+d_2(x-d_1,y)-d_3]\\ &-\lambda\Big\{\frac{\partial}{\partial x}\Big[\frac{d_{3,x}}{(1+I_{l,x}^2)^2}\Big]+\frac{\partial}{\partial y}\Big[\frac{d_{3,y}}{(1+I_{l,y}^2)^2}\Big]\Big\}.\end{aligned} \tag{7}$$

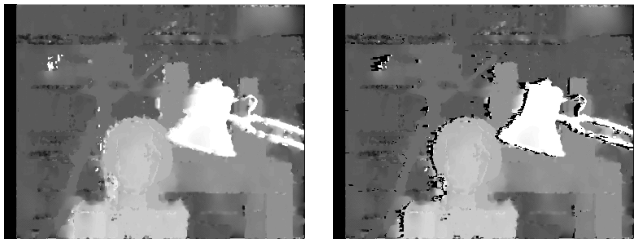
Along with the refinement process of $d_1(x,y)$ and d_3 (x,y), $d_2(x,y)$ also needs to be updated. This was done by solving the Euler-Lagrange equation of $d_2(x,y)$ associated with the data fidelity and regularization functionals between I_{r1} and I_{r2}:

$$\begin{aligned}\frac{\partial d_2}{\partial t}=&[I_{r1}(x,y)-I_{r2}(x-d_2,y)]\times I_{r2,x}(x-d_2,y)\\ &-\lambda\Big\{\frac{\partial}{\partial x}\Big[\frac{d_{2,x}}{(1+I_{r1,x}^2)^2}\Big]+\frac{\partial}{\partial y}\Big[\frac{d_{2,y}}{(1+I_{r1,y}^2)^2}\Big]\Big\}.\end{aligned} \tag{8}$$

Thus, (6), (7) and (8) form a set of coupled PDEs, and the refinement of $d_1(x,y)$ is achieved via (6) with the updates of $d_2(x,y)$ and $d_3(x,y)$ from (8) and (7). In the refinement process, the disparity values for the feature points previously identified are kept unchanged, since these values from previous estimation by phase methods already have high reliability. The reason for this is because that we found in our previous experiments, as well as in other literature, that with the increasing number of iterations in solving the PDEs, the object contours would get more and more blurred, and some thin objects like the handle of the lamp in *Tsukuba* as well as the twigs in *Flower Garden* could get faded out. From the final results of refinement, as shown in Fig. 4(a) for *Tsukuba* and Fig. 5(a) for *Flower Garden*, we find that the object boundaries are well preserved, while those noisy areas in coarse disparity maps get smoothed. The number of iterations we used is 2000. Also, by comparing with our other refinements involving only one pair of stereo images, as shown in Fig. 6, the refinements using three images can give perceivable improvement in some areas with object contours.

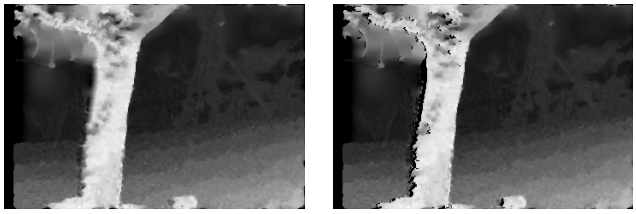
4. OCCLUSION DETECTION

The occlusion mask $o_1(x,y)$ for $d_1(x,y)$ was obtained from the fact that, for disparity of the left image, occlusions oc-



(a) (b)

Fig. 4. (a) Refinement for *Tsukuba*. (b) Disparity map with occlusion (in black).



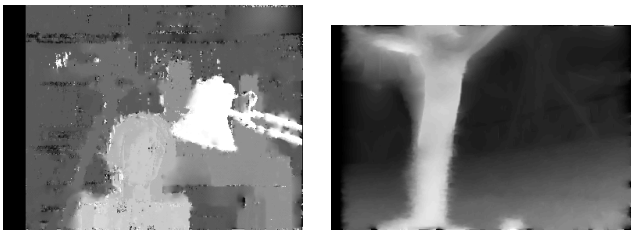
(a) (b)

Fig. 5. (a) Refinement for *Flower Garden*. (b) Disparity map with occlusion (in black).

cur along the left side of an object. Thus we can search the area where the horizontal disparity derivative is larger than a threshold. Once we found that the magnitude of the x-derivative of the intensity for a pixel which lies to the right of the detected disparity derivative is larger than another threshold, then this pixel will be determined as an object boundary, and the pixels to the left of this boundary pixel will be labeled as occluded pixels. The number of occluded pixels to be labeled is determined by the value of the disparity difference where the horizontal disparity derivative being larger than a threshold occurs. The results of $d_1(x, y)$ with $o_1(x, y)$ are shown in Fig. 4(b) for *Tsukuba* and Fig. 5(b) for *Flower Garden*.

5. CONCLUSION AND FUTURE WORK

We have developed a new algorithm to increase the precision for disparity estimation in a multi-baseline variational regularization framework, as well as an effective occlusion detection scheme. The experimental results show that our algorithm is robust for images with different features. From the visual effects of the disparity maps with occlusion detected, our algorithm appears to be helpful for both 3D model construction from stereo images, and view interpolation. Since for disparity estimation, especially with occlusion, there is no standard way to give a quantitative performance measure, in this paper, we only judge the improvement that our new algorithm brings by visual effects in the disparity maps with occlusion. In the future, we will further develop a refinement functional involving occlusion masks, and apply the algorithm to multi-baseline stereo with more images, so that novel views can be rendered and compared to existing images to obtain a final performance measure.



(a) (b)

Fig. 6. Refinement using only one stereo pair: (a) *Tsukuba*. (b) *Flower Garden*.

6. ACKNOWLEDGEMENTS

This work was supported by the Natural Sciences and Engineering Research Council of Canada under the research network *LORNET*.

7. REFERENCES

[1] L. Alvarez, R. Deriche, J. Sanchez, and J. Weickert, "Dense disparity map estimation respecting image discontinuities: A PDE and scale-space based approach," *Journal of Visual Communication and Image Representation*, vol. 13, pp. 3–21, Jan. 2002.

[2] C. Strecha, T. Tuytelaars, and L. Van Gool, "Dense matching of multiple wide-baseline views," in *Proc. IEEE Int. Conf. Computer Vision*, 2003, vol. 2, pp. 1194–1201.

[3] D.J. Fleet, "Disparity from local weighted phase-correlation," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, 1994, pp. 48–56.

[4] M. Okutomi and T. Kanade, "A multi-baseline stereo," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, pp. 353–363, 1993.

[5] S.B. Kang, R. Szeliski, and J. Chai, "Handling occlusion in dense multi-view stereo," in *Proc. IEEE Conf. Computer Vision Pattern Recognition*, Dec. 2001, vol. 1, pp. 103–110.

[6] X. Huang and E. Dubois, "Disparity estimation for the intermediate view interpolation of stereoscopic images," in *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, 2005, vol. 2, pp. 881–884.