

DISPARITY ESTIMATION FOR THE INTERMEDIATE VIEW INTERPOLATION OF STEREOSCOPIC IMAGES

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

This paper presents a disparity estimation method that combines the phase-based disparity estimation algorithm with an energy-based variational refinement approach, and adapts the method for the application of intermediate view interpolation. The Gabor transform is implemented using a set of quadrature-pair filters for finding the local phase information as well as for estimating the correspondences in stereo pairs. Then the estimated coarse disparity is applied to a variational refinement process which involves solving a partial differential equation (PDE) associated with an edge-preserving regularization functional. Two stereo image pairs in which the original intermediate images are available were used to test the algorithm, where the interpolated intermediate images based on our method could be compared with the original ones. The performance with PSNR results shows that our method can provide very good images without introducing obvious artifacts.

1. INTRODUCTION

Stereoscopic viewing devices (e.g. shutter glasses) provide an effective way for perceiving the depth information in real-world scenes. However, the head movement of the viewer leads to an feeling in which a static object will seem to rotate. This is caused by the motion-parallax conflict because the screen parallax does not change with the motion of the viewer. To alleviate such conflict, intermediate views of stereoscopic images are needed so that the displayed images can be adjusted along with the head movement of the viewer.

For the interpolation of intermediate views, the correspondence information or the disparity field between the stereo images is required. Usually the algorithms for disparity estimation can be classified into *the area-based methods*, *the energy-based methods* and *the phase-based methods*. The area-based methods are the most often used approaches; they find the correspondence for a block area in

one image by comparing it with shifted blocks in a search region in the other image. They perform well in relatively textured areas, but will cause problems in slanted and untextured areas, and are also sensitive to brightness variations. The energy-based methods estimate the disparity using a minimization and regularization formulation [1][2], which usually consists of an iterative solution of the associated discretized partial differential equation (PDE). These methods can estimate the disparity with high accuracy, under the condition that the coarse disparity that is used as the initial value for the associated PDE has enough accuracy; otherwise the iterative process in solving the PDE is very likely to fall into local minima. Usually the results from the area-based methods are used as the initial values for the energy-based approach. The phase-based methods use the local phase information of the images to find the disparity [3][4]. They are more robust to different texture levels and slanted surfaces, less sensitive to brightness variations, and thus could give better results than the area-based methods. We use the phase-based methods for a more accurate initial disparity, and then use the energy-based approach to refine that coarse disparity.

Since our final goal is to interpolate an intermediate image, we need to modify the formulation and the matching process of the phase-based methods to fit this purpose. This will be shown in section 2. As for our knowledge, there is no previous work with this respect. Besides, the data-fidelity term in the refinement functional will also be changed accordingly, which will be shown in section 3. The conclusion will be given in section 4.

2. DISPARITY ESTIMATION USING GABOR TRANSFORM

The Gabor functions are Gaussian functions modulated by complex sinusoids. For the 2-D case, they are defined as

follows:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[-\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) / 2 \right] \quad (1)$$

$$\times \exp [2\pi j (\omega_{10}x + \omega_{20}y)]$$

where ω_{10} and ω_{20} define the spatial frequencies in x and y directions respectively. Its Fourier transform has the form:

$$G(\omega_1, \omega_2) = e^{\{-2\pi^2[\sigma_x^2(\omega_1 - \omega_{10})^2 + \sigma_y^2(\omega_2 - \omega_{20})^2]\}}. \quad (2)$$

In order to show how the disparity can be obtained from the Fourier phase information of the stereo images, first assume that the right image is a pure horizontal translation of the left image $I_r(x, y) = I_l(x - d, y)$, where d is constant over the whole image. Its related Fourier transform is $\hat{I}_r(\omega_1, \omega_2) = \hat{I}_l(\omega_1, \omega_2)e^{-j\omega_1 d}$. Therefore, we have:

$$\frac{\hat{I}_l(\omega_1, \omega_2)\hat{I}_r^*(\omega_1, \omega_2)}{|\hat{I}_l(\omega_1, \omega_2)||\hat{I}_r(\omega_1, \omega_2)|} = e^{j\omega_1 d}. \quad (3)$$

Hence from the above normalized phase-correlation we can obtain the phase difference between $\hat{I}_l(\omega_1, \omega_2)$ and $\hat{I}_r(\omega_1, \omega_2)$, and the disparity can be obtained by taking the inverse Fourier transform of the correlation product, resulting in an impulse at the location d .

In practice, the disparity values vary over the whole image. Thus it is desirable to measure the phase difference locally rather than globally. In order to do this we need to use the windowed Fourier transform. The best choice for the window is to use the Gabor function (1) because the Gaussian window performs the localization in both the spatial and the frequency domains simultaneously, as can be seen from (1) and (2). The localization in the spatial domain limits the disparity in a narrow neighborhood while in the frequency domain it allows to calculate the disparity adequately.

The method that we used in this paper to estimate the disparity from the local phase correlation using the Gabor transform is mainly based on the idea in [4], in which a set of quadrature-pair Gabor filters are used. Each quadrature-pair Gabor filter is a set of discretized samples of (1) with different ω_{10} and ω_{20} , and is used for the filtering of the stereo images to obtain the approximate Gabor transform coefficients at those ω_{10} and ω_{20} . Assume that the outputs of k^{th} filter pair are $G_l^k(x, y)$ and $G_r^k(x, y)$ for the left and right images respectively. In [4], instead of taking the normalized product of the left and right outputs directly like (3), a similar product of the left output with a preshifted right output is taken:

$$C^k(x, y, d) = \frac{G_l^k(x, y)G_r^{k*}(x - d, y)}{|G_l^k(x, y)||G_r^k(x - d, y)|} \quad (4)$$

where d is the shifting parameter. Then the results of (4) are summed together for all the filter pairs as $S(x, y, d) =$

$\sum_k C^k(x, y, d)$. The disparity for a position (x, y) is determined to be the d which gives a peak for the real part of $S(x, y, d)$ at (x, y) and a minimum value for its imaginary part.

Since we are dealing with intermediate view interpolation, the disparity values we need to estimate should fit the purpose of finding the suitable pixel values for the interpolated image at a specific intermediate location, rather than finding the disparity between the left and the right images directly [5]. To make this clear, assume the distance between I_l and I_r is normalized, and the position of the intermediate image is at the normalized distance $0 \leq \alpha \leq 1$ from I_l and therefore $1 - \alpha$ from I_r . For a pixel of the intermediate image at coordinates (x, y) , we need to find a disparity d that gives a best match between $I_l(x + \alpha d, y)$ and $I_r(x - (1 - \alpha)d, y)$. This is different from finding a match between $I_l(x, y)$ and $I_r(x - d, y)$ for the disparity of the left image at coordinates (x, y) . Therefore, we need to modify (4) as follows:

$$C_I^k(x, y, d) = \frac{G_l^k(x + \alpha d, y)G_r^{k*}(x - (1 - \alpha)d, y)}{|G_l^k(x + \alpha d, y)||G_r^k(x - (1 - \alpha)d, y)|}. \quad (5)$$

Then the disparity value d for (x, y) will be estimated by summing $C_I^k(x, y, d)$ from all filters as $S_I(x, y, d) = \sum_k C_I^k(x, y, d)$ and looking for the d that gives a maximum value in the real part of $S_I(x, y, d)$ and a minimum value in its imaginary part. After obtaining the disparity field, the intermediate image I_I at the normalized position α will be interpolated as [5]:

$$I_I(x, y) = (1 - \alpha)I_l(x + \alpha d, y) + \alpha I_r(x - (1 - \alpha)d, y). \quad (6)$$

We use three values $\{\pi/16, \pi/8, \pi/4\}$ for the central frequency $\omega_0 = \sqrt{\omega_{10}^2 + \omega_{20}^2}$ of the Gabor filters. For each frequency, there are four filter pairs tuned to orientations $0^\circ, 45^\circ, 90^\circ$ and 135° respectively. To test the performance of our method quantitatively, we use 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 frame 2 – 4 from *Tsukuba* and frame 21 – 23 from *Flower Garden*, with the first frame as left image and the last frame as right image for each of the sequence. We interpolate an intermediate image at the middle position ($\alpha = 0.5$), and calculate the peak signal-to-noise ratio (PSNR) between the interpolated image and the original middle frame image, for each sequence. The original middle images of *Tsukuba* and *Flower Garden* are shown in Fig. 1(a) and Fig. 2(a) respectively. The estimated disparity for each of them are shown in Fig. 1(b) and Fig. 2(b) respectively. The PSNR of the interpolated middle image is 27.95dB for *Tsukuba*, and 26.26dB for *Flower Garden*.



Fig. 1. (a) Original middle image of *Tsukuba*; (b) the coarse disparity estimated by the phase method



Fig. 2. (a) Original middle image of *Flower Garden*; (b) the coarse disparity estimated by the phase method

3. DISPARITY REFINEMENT

From a coarse point of view, the disparity estimated by the phase-based method is good. However, there are still some obvious errors in the disparity map, like some noisy values or abrupt changes in a flat area, or some distorted edges, as can be seen from Fig. 1(b) and Fig. 2(b). To alleviate such errors we need to get the disparity values for continuous surfaces changing smoothly, while maintaining the disparity discontinuities at the object boundaries. To achieve such properties, the energy-based variational regularization can be applied as a refinement process for the disparity estimation. The energy functional used to estimate the disparity $d(x, y)$ usually contains two terms: $E(d) = E_D(d) + \lambda E_S(d)$, where $E_D(d)$ is a data fidelity term which is defined in our case as

$$E_D(d) = \iint [I_l(x + \alpha d, y) - I_r(x - (1 - \alpha)d, y)]^2 dx dy, \quad (7)$$

and $E_S(d)$ is a regularization term which controls the smoothing of the disparity field, together with the regularization coefficient λ .

3.1. Regularization Functional

The variational regularization has had extensive application in optical flow estimation, as well as some for disparity estimation [1][2]. The idea in [1] 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 get smoothed, and 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 is mainly based on the functional in [1]:

$$E_S(d) = \iint [(\nabla d)^t D(\nabla I_l) \nabla d] dx dy \quad (8)$$

where $D(\nabla I_l)$ is a matrix defined by:

$$D(\nabla I_l) = \frac{1}{|\nabla I_l|^2 + 2v^2} \left\{ \begin{bmatrix} \frac{\partial I_l}{\partial y} \\ -\frac{\partial I_l}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial I_l}{\partial y} \\ -\frac{\partial I_l}{\partial x} \end{bmatrix}^t + v^2 Id \right\} \quad (9)$$

where Id is the identity matrix and v is an arbitrary positive real number. The smoothing term $E_S(d)$ is anisotropic: in homogeneous areas the disparities are smoothed in all directions, while in textured areas including edges the smoothing is mainly along the edge but not across it.

To minimize the energy $E(d)$, we apply a gradient descent method to solve its associated Euler-Lagrange equation:

$$\begin{aligned} \frac{\partial d}{\partial t} = & [I_r(x - (1 - \alpha)d, y) - I_l(x + \alpha d, y)] \\ & \times [\alpha I_{lx}(x + \alpha d, y) + (1 - \alpha)I_{rx}(x - (1 - \alpha)d, y)] \\ & + \lambda \text{div}(D(\nabla I_l) \nabla d). \end{aligned} \quad (10)$$

The numerical implementation of the above PDE is given by the forward Euler method, and the spatial derivatives are calculated by the central difference scheme. We refer to [1] for the implementation details.

3.2. Simulation Results

The parameters we used in solving (10) iteratively are $\lambda = 2.5$, time step 0.05, iterations 800 – 1000. The initial disparity values are from the results of our phase-based methods. The refinement results for *Tsukuba* and *Flower Garden* are shown in Fig. 3(a) and Fig. 4(a) respectively.



Fig. 3. (a) Refined disparity field for *Tsukuba*; (b) interpolated middle image, 30.04dB

Comparing Fig. 3(a) and Fig. 4(a) with Fig. 1(b) and Fig. 2(b) we can see that the noise in the coarse disparity



Fig. 4. (a) Refined disparity field *Flower Garden*; (b) interpolated middle image, 26.93dB

maps are mostly removed, and the disparities on the object surfaces are somewhat smoothed. However, we also found that if more iterations are applied to (10), the object contours begin to get blurred, e.g. the contours of the lamp in *Tsukuba*. This is not unexpected since in our regularization functional (8), the smoothing term is quadratic with respect to ∇d . It is well known that a quadratic regularization functional, like the Tikhonov functional in image restoration or the Horn and Schunck functional in optical flow estimation, will unsharp the edges of the image. Although the functional (8) is derived in [1] from the Nagel and Enkelmann functional, which is shown to be the best quadratic smoothness constraint for optical flow estimation [6], it is known that for image restoration the total variation regularization will do better than quadratic terms in keeping the edges sharp and clear. However, we could not combine it with the functions that control the disparity smoothing, because in (8), if we use a term like $D(\nabla I_l)|\nabla d|$, then a new term $|\nabla d|$ in the denominator of its Euler-Lagrange equation will break the smoothing controlled by the function $D(\nabla I_l)$.

The PSNR values of the interpolated image are 30.04dB for *Tsukuba*, and 26.93dB for *Flower Garden*. Comparing the PSNR values with the ones before the refinement, we can see that the PSNR of *Tsukuba* increased more than 2dB, while for *Flower Garden* it is 0.67dB. This could be explained by the fact that *Flower Garden* contains lots of highly featured areas, while for *Tsukuba* the featured areas are not very dense. Thus the smoothing of the regularization functional with a quadratic term will give more benefits for *Tsukuba*.

4. CONCLUSION AND FUTURE WORK

In this paper, we combined the phase-based method and the variational regularization to estimate the disparity, and adapted these algorithms for the application of intermediate view interpolation. The experimental results show that our method is effective, and is robust in the sense that it can be applied to images with different features, although the refinement could give more gain to images with moderate feature levels. From both the visual effects of the dispar-

ity map, as well as the interpolation performance, our algorithm appears to be helpful for both the image-based interpolation/rendering and the 3D model construction from stereo images.

In the future, we will try to apply some combinatorial optimization algorithms like the graph cuts [7] to our phase-based method so that an energy functional including phase coefficients as well as total variational regularization can be discretized directly and optimized using graph cuts.

5. ACKNOWLEDGEMENTS

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

6. 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] L. Robert and R. Deriche, "Dense depth map reconstruction: A minimization and regularization approach which preserves discontinuities," in *Proceedings of the 4th European Conference on Computer Vision*, Apr. 1996.
- [3] D.J. Fleet, Jepson, and M. Jenkin, "Phase-based disparity measurement," *CVGIP: Image Understanding*, vol. 53, pp. 198–210, 1991.
- [4] D.J. Fleet, "Disparity from local weighted phase-correlation," in *IEEE International Conference on Systems, Man, and Cybernetics*, 1994, pp. 48–56.
- [5] A. Mancini and J. Konrad, "Robust quadtree-based disparity estimation for the reconstruction of intermediate stereoscopic images," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, 1998, vol. 3295, pp. 53–64.
- [6] M.A. Snyder, "On the mathematical foundations of smoothness constraints for the determination of optical flow and for surface reconstruction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1105–1114, 1991.
- [7] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. IEEE Int. Conf. Computer Vision*, July 2001, vol. 2, pp. 508–515.