Disparity estimation and occlusion detection in aerial scenes based on the relationship between phase-based and area-based stereo

Yi Xu (徐 奕), Jun Zhou (周 军), and Yuanhua Zhou (周源华)

Institute of Image Communication & Information Processing, Shanghai Jiaotong University, Shanghai 200030

Received March 7, 2003

In this letter, it is shown that there exists relationship between phase-based and area-based stereo in spite of their different motivations. A new cost function is defined based on this clue and an improved cost-minimization framework is presented. It is suitable for disparity estimation and occlusion detection in aerial scenes.

OCIS code: 330.1400.

Phase-based stereo has generated great interests mainly because the phase behavior in band-pass filtered versions of different views of a 3D scene encodes the original image structure information and holds its stability to geometric deformations[1]. But despite these merits, phase singularities result in mismatches in phase-based stereo. Phase singularities are generally removed depending on some criterions and these holes in the disparity map are often filled using smoothness constraints[2,3] after matching is completed.

By quantizing a view into a number of subregions or blocks, it is possible to apply an area based similarity metric to find the most likely correspondence between the same regions from different views[4]. In area-based stereo, intensity value at each point is regarded as a feature and correlation metric is dominantly used to get corresponding blocks. However, due to sampling effects and illumination variations, intensity is not always the same at corresponding points.

Aerial imagery is one of the standard data sources for the acquisition of topographic objects. Image deformations typically exist between aerial stereo pairs due to the time interval between them. Moreover, the narrow occluding objects rarely appear in aerial stereo pairs for the width of the baseline can be ignored compared with the distance from camera to objects.

In this letter, the relationship between phase-based stereo and area-based stereo is discussed and a new cost function is proposed to offer reliable similarity measure with regard to phase singularities and image deformations. The minimization of the cost function is performed with the help of dynamic programming (DP) technique to achieve high performance evaluation. Furthermore, highly-reliable matches (HRMs) and occlusion models in the 2D matching space ensure reliable occlusion detection obtained simultaneously with disparity estimations.

Phase-based stereo methods can be categorized into two groups–phase-difference methods and phase-correlation methods. In phase-difference methods, the local image disparity at a specific position x, for an initial guess d_0, is defined to be the shift $d(x)$ such that

$$\phi_l\left[x-\frac{d(x)}{2}\right]=\phi_r\left[x+\frac{d(x)}{2}\right], \tag{1}$$

and $|d(x)-d_0|$ is as small as possible[2]. With linear phase model, the corresponding disparity estimator takes the form

$$d(x)=\frac{[\phi_l(x)-\phi_r(x)]_{2\pi}}{0.5[\phi_l'(x)+\phi_r'(x)]}, \tag{2}$$

where $[\psi]_{2\pi}$ denotes the principal part of ψ that lies between $-\pi$ and π.

Phase-correlation methods use Fourier phase for signal registration[5]. The methods are often derived assuming pure translation between two images

$$I_r(x)=I_l[x-d(x)]. \tag{3}$$

Let $\hat{i}_l(\omega)$, $\hat{i}_r(\omega)$ respectively represent the Fourier transform of left and right images, and $\rho_l(\omega)$, $\rho_r(\omega)$ respectively denote the amplitude spectra of left and right images, phase-correlation methods measure disparity by finding peaks in

$$F^{-1}\left[\frac{\hat{I}_l(\omega)\hat{I}_r^*(\omega)}{\rho_l(\omega)\rho_r(\omega)}\right], \tag{4}$$

where F^{-1} denotes the inverse Fourier transform and '*' denotes the conjugate counterpart.

From above analysis, it can be deduced that phase-based stereo is derived from the principle that computing the disparity amounts to determining the shift between the left and right views so that the phase structures of them become equal. In area-based stereo, cross-correlation technique is dominantly exploited. So in what follows, it is discussed the relationship between cross-correlation area-based stereo and phase-based stereo.

To localize isomorphic phase structures in two views, the solution of disparity corresponds to selecting d to maximize the function

$$\begin{aligned} f(d)&=\frac{F^{-1}[\hat{I}_r(\omega)\hat{I}_l^*(\omega)]}{[\int_{-\infty}^{\infty}\rho_r^2(\omega)\mathrm{d}\omega\int_{-\infty}^{\infty}\rho_l^2(\omega)\mathrm{d}\omega]^{1/2}}\\ &=\frac{\frac{1}{2\pi}\int_{-\infty}^{\infty}\hat{I}_r(\omega)\hat{I}_l^*(\omega)e^{j\omega d}\mathrm{d}\omega}{[\int_{-\infty}^{\infty}\rho_r^2(\omega)\mathrm{d}\omega\int_{-\infty}^{\infty}\rho_l^2(\omega)\mathrm{d}\omega]^{1/2}}. \end{aligned} \tag{5}$$

The maximization amounts to finding the 'phase correction' term ωd which maximizes the normalized inner product between the spectra of the binocular images. According to Parselval's theorem, Eq. (5) can be written in terms of the spatial domain as

$$f(d) = \frac{\int_{-\infty}^{\infty} I_l(x) I_r(x+d)\mathrm{d}x}{[\int_{-\infty}^{\infty} I_r^2(x)\mathrm{d}x \int_{-\infty}^{\infty} I_r^2(x)\mathrm{d}x]^{1/2}}, \quad (6)$$

which is just the normalized cross correlation between $I_l(x)$ and $I_r(x)$, usually the algorithmic model in area-based methods.

As mentioned above, it could be argued in principle that the solution of disparity at which the phase structures of different perspectives become equal in frequency domain amounts to maximizing the correlation coefficient of them in spatial domain.

Minimization of a cost function over some local region is a well-known strategy for computing dense correspondences. This poses the question of selecting a cost function as to what is the most appropriate way of setting similarity measures. To reduce the complexity of disparity measurement, the stereo matching problem mentioned here is simplified to an intra-scanline pixel matching problem. Given the disparity d and the point at location x on the scanline n in the reference (left) view, the cost function of our approach is defined as

$$\mathrm{cost}(x,d,n) = \frac{1}{w \times w} \sum_{i,j=-w/2}^{w/2} [\gamma \times \rho_{i,j}^{\mathrm{norm}} \times \Delta\phi_{i,j} + (1 - \rho_{i,j}^{\mathrm{norm}}) \times \Delta G_{i,j}], \quad (7)$$

where

$$\Delta\phi_{i,j} = |[\phi_l(x+i, n+j) - \phi_r(x+i+d, n+j)]_{2\pi}|, \quad (8)$$

$$\rho_{i,j}^{\mathrm{norm}} = [\rho_l(x+i, n+j) \times \rho_r(x+i+d, n+j) + \varepsilon] / \{ \sum_{i,j=-w/2}^{w/2} [\rho_l(x+i, n+j) \times \rho_r(x+i+d, n+j) + \varepsilon] \}, \quad 0 < \varepsilon < 1, \quad (9)$$

and

$$\Delta G_{i,j} = |\frac{1}{a} \times [I_l(x+i+1, n+j) - I_l(x+i-1, n+j)] - [I_r(x+i+d+1, n+j) - I_r(x+i+d-1, n+j)]|. \quad (10)$$

where w is the size of the matching window, ϕ, ρ and I denote the phase, amplitude and intensity value at specified locations in the matching window. Coefficient γ is imposed on the cost function to hold $\Delta\phi_{i,j}$ and $\Delta G_{i,j}$ in the same variation range. ε in Eq. (9) is used to hold the denominator not equal to zero. Coefficient a in Eq. (10) reflects the linear illumination distortions between two views.

Since phase structures should be invariant at corresponding points, we can use phase differences to describe the similarity measure. However, phases are unreliable near singularities in space-time where the filter output passes through the origin in the complex plane. Given a point in the reference view and its corresponding point in the other view, if the phase of either point is near singularities, the amplitude product must be very small. In our cost function, $\Delta\phi_{i,j}$ is weighted with normalized amplitude product $\rho_{i,j}^{\mathrm{norm}}$ to suppress the ill effects of the singularities on the cost evaluation aggregated by the phase differences over a local support. It works in the case when the singularities occur only at some isolated points in the matching window. While the worst case occurs that the phases of the points in the matching window are all singularities, we need the correlativity between two points in space domain to reflect how much they match in frequency domain. The costs aggregated by the term $\Delta G_{i,j}$ just measure the correlativity between two regions. It is notified that local phase data is preferred to correlation coefficient in application to stereo matching with respect to image deformations, such as scale changes up to 20%, noise and common illumination variations. So phase information is emphasized in the cost evaluation while the intensity information is subdued with local energy increasing.

The above analysis demonstrates that stability constraints are implicit in the cost function. The accuracy improvement is therefore achieved without additional detection and removal of singularities and the cost evaluation is robust to image deformations.

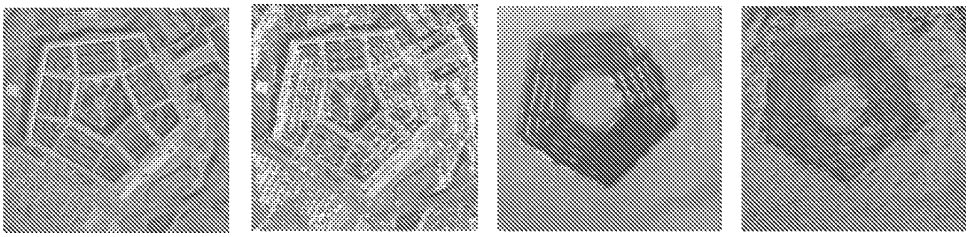
Given cost function, the disparity measurement is just to minimize the sum of the costs along each epipolar. Occlusion is a critical and difficult phenomenon to be dealt with in the minimization process. Without respect to narrow occluding objects, the quantitative measures which specifically measure occlusion detection indicate that DP has a low probability of false alarm, a high probability of correct detection, a low mean-squared error and effectively keeps local disparities near each other without using a regularization framework or a relaxation technique, thus achieving fast minimization[6]. The computation of the cost matrix for DP is carried out in the 2D matching space which could incorporate occlusion models directly into the disparity estimation process[7]. However, the correct guidance of HRMs is crucial for the solution path.

A pixel-to-pixel feature matching process is presented here to get HRMs and only a little computation complexity is incurred. The amplitude ρ of the filtered output at each point is used as the local feature. To reduce ambiguity problems, points with higher amplitude have a higher probability of being selected. Given a feature at the reference location x_l on scanline n in the left view, the disparity value d is computed using phase-difference method, as shown in Eq. (12). If the corresponding point is at location x_r in the other view, there should be

$$\rho_l(x_l) > \mathrm{th}_\rho > 0 \text{ and } \rho_r(x_r) > \mathrm{th}_\rho > 0 \text{ and } |x_r - (x_l + d)| < \mathrm{th}_d, \quad (11)$$

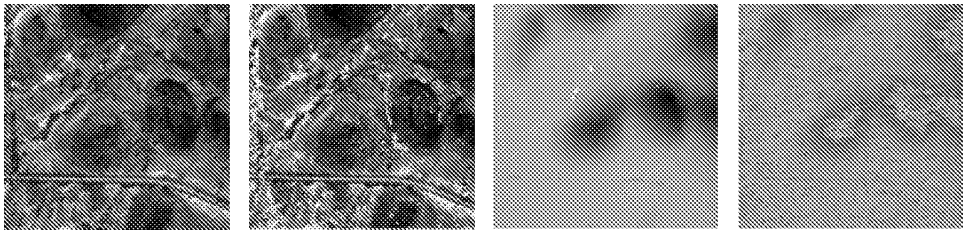
where

$$d = \frac{\phi_l(x_l) - \phi_r(x_l + d_0)}{0.5 \times [\phi_l'(x_l) + \phi_r'(x_l + d_0)]} + d_0, \quad (12)$$



(a) (b) (c) (d)

Fig. 1. Experiment results with 'pentagon' stereo images. (a) Left image; (b) HRMs in the left image; (c) computed disparity map with the proposed approach, where white regions are occluded in the right view; (d) computed disparity map with phase-difference method.



(a) (b) (c) (d)

Fig. 2. Experiment results with 'hill' stereo images. (a) Left image; (b) HRMs in the left image; (c) computed disparity map with the proposed approach, where white regions are occluded in the right view; (d) computed disparity map with phase-difference method.

d_0 is the initial disparity, th_ρ and th_d are thresholds. For each scanline, the minimal-cost traversal through the HRMs in the 2D matching space, determines the disparities and occlusions simultaneously and significantly reduces the algorithm's occlusion-cost sensitivity and computation complexity.

The wavelength of the complex filter is equal to 2.044. In Figs. 1 and 2, two 512 × 512 aerial image pairs are selected to demonstrate the validity of the proposed approach. If there are areas occluded in the right view, they are marked in the disparity map with the highest gray value 255. The HRMs are marked with white crosses in stereo images. To demonstrate that further processing is necessary in the conventional phase-difference approach to obtain results comparable to those of our approach, phase singularities are not detected and removed in both approaches. Here some parameters in our approach are given. $w = 5, \varepsilon = 0.5, \mathrm{th}_\rho = 20$, $\mathrm{th}_d = 0.5$. The occlusion penalty cost is assigned 30 with γ equal to 10.

The advantages of the proposed approach can be considered from its application to aerial scenes, its improvement of recent phase-based stereo methods and the computation complexity. The cost function provides reliable similarity measures with regard to image deformations in aerial images. The stability constraints implicit in the proposed cost function render the algorithmic model more self-adaptive, not requiring criterions and empirical thresholds to deal with singularity problem. The aggregation of costs over a local support automatically imposes smoothness constraints on the disparity map. In our approach, the measurement of disparity values, the accuracy improvement related to singularity problem and the detection of occlusions are all integrated in a cost-minimization framework, rather than regarded as individual parts of a stereo matching process. With the help of DP search process and correct guidance of HRMs, high performance evaluation is achieved. The matching process only needs tens of seconds on PIII 1G PC.

This work was supported by the National Natural Science Foundation of China under Grant No. 69905003. The authors' e-mail addresses are gradually nice @hotmail.com (Y. Xu), zhoujun@sjtu.edu.cn (J. Zhou), and yhzhou@cdtv.org.cn (Y. Zhou).

References

1. D. J. Fleet and A. D. Jepson, IEEE Trans. PAMI **15,** 1253 (1993).
2. D. J. Fleet, A. D. Jepson, and M. R. M. Jenkin, CVGIP: Image Understanding **53,** 198 (1991)).
3. M. H. Ouali, D. Ziou, and C. Laurgeau, in *Proceedings of Vision Interface'99* 439 (1999)).
4. R. A. Lane and N. A. Thacker, http://citeseer.nj.nec.com /lane96stereo.html (1996).
5. D. J. Fleet, in *Proceedings of IEEE International Conference on System, Man and Cybernetics* 48 (1994).
6. G. Fielding and M. Kam, Pattern Recognition **33,** 1511 (2000).
7. C. J. Tsai and A. K. Katsaggelos, IEEE Trans. on Multimedia **1,** 18 (1999).