COOPERATIVE STEREO MATCHING USING QUATERNION WAVLETS AND TOP-DOWN SEGMENTATION

Yi Xu[1], Xiaokang Yang[1], Peifeng Zhang[2], Li Song[1] and Leonardo Traversoni[3]
Institute of Image Communication and Information Processing
Shanghai Jiao Tong University, Shanghai, PRC 200240
Univ. Autonoma Met, Iztapalapa
[1]{xuyi,xkyang,song_li}@sjtu.edu.cn [2]zpfxellose@gmail.com [3]ltd@xanum.uam.mx

ABSTRACT

We explore the principles of quaternion wavelet construction for achieving multiscale analysis of geometric image features. Then the quaternion wavelets are applied to propose a cooperative stereo matching algorithm using top-down segmentation-based disparity propagation. Without bidirectional matching to remove ambiguous outliers, uniqueness constraint is enforced on cost function by inhibiting the matches along similar sightlines. To produce smooth disparity maps with the discontinuities well-preserved, cost aggregation is performed in segmentation-based local support and high confidence matches serve as heavyweight seeds for disparity propagation in the supports. Compared with the current matching methods based on quaternion wavelets, the main merit of the proposed algorithm is that the matching results are encouraging in extensive comparison data, ranging from calibrated images to uncalibrated images, indoor images to aerial images.

1 INTRODUCTION

Quaternion wavelet transform (QWT) achieves much attention in recent years as a new multiscale analysis tool for geometric image features. It is an extension of the real wavelet transform and complex wavelet transform (CWT) by using the quaternion algebra and the 2-D Hilbert transform of filter theory [1][2]. Its dominant applications in the existing research work are involved in color image processing [3], object recognition [4], optical flow estimation [1][2] and stereo matching [5][6]. The authors pay much attention to the differences between QWT, CWT and discrete wavelet transform (DWT) [2][5] and quaternion Gabors are the most common-used filters [1][4]. A few authors use the tensor products of Kingsbury's dual-tree complex wavelets [7] for analytic signal generation [2][5]. The principles of quaternion wavelet construction are somewhat fragmentary in the above works. It would be investigated in this paper and arranged as the main content in section 2.

We are also interested in the issue of stereo matching using quaternion wavelets. The 2-D phase concept of QWT is attracting for uncalibrated matching tasks because it encodes the relative location of intrinsically 2-D image features (corner-like). Thus epipolar geometry need not be recovered prior to the matching process. Bülow is the first foregoer introducing quaternion wavelets into 2-D disparity computation [1]. Since only single-scale phase information is exploited in Bülow's research, the disparity search range is limited to the wavelength of the used quaternion Gabor filter. Later, Corrochano [2] and Chan [5] et al. exploited multiscale phase unwrapping technique to enlarge the disparity search scope and the density of disparity is improved through multiscale phase representation. However, the negative effects of phase singularity may deteriorate the matching results through the disparity error propagation from coarse scales to fine scales. In our previous work in [6], we introduced the disparity coherent constraint and phase stability constraint to greatly suppress the contribution of phase singularity to disparity estimation. The experimental results demonstrate the potential of QWT for practical use in an uncalibrated stereo rig. Still, there are some problems left in the matching algorithm of [6]. Due to the fixed support window for cost aggregation, the discontinuity of disparity is not well preserved and ambiguous matches increase in large textureless regions. In addition, the uniqueness constraint is enforced on the reference view unidirectionally, thus the best disparity is selected without inhibiting the matches along similar sightlines in the other view.

In section 3, we propose a cooperative stereo matching scheme as an improved version of the work in [6]. The matching scheme is designed to achieve following developments: (1) Find trade-off in disparity map between smoothness and precise boundary using segment-based supports. (2) Reduce the ambiguous matches via cross validation without bidirectional matching. (3) Improve the accuracy by robust outlier rejection during cost aggregation.

In section 4, we testify our matching scheme with various stereo pairs, ranging from calibrated images to uncalibrated images, indoor images to aerial images, including the ones with ground truth for quantitative evaluation. In section 5, we conclude the paper.

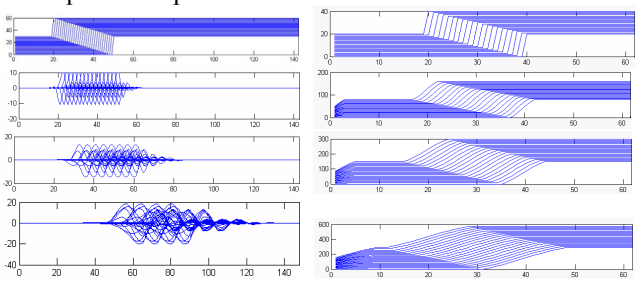
2 PRINCIPLES OF QUATERNION WAVELET CONSTRUCTION

2.1 Linear-phase and shift-invariance property

It is noted that the linear-phase property of quaternion

wavelets is important for multiscale signal analysis, thus no phase compensation is needed in multiscale signal decomposition. Biorthogonal wavelet bases are symmetric or asymmetric, which guarantee the linear phase property. Whilet, this property can't be found in ordinary orthogonal wavelet bases except Harr wavelet. Besides the most common quaternion Gabors, some authors exploit tensor products of linear-phase complex wavelets to build quaternion wavelets, where the complex wavelet is generated from biorthogonal wavelets [5][6].

To capture geometric image features nonoscillatorily, one important property of the filter is that a shift in the time domain should cause no change in the magnitude spectrum. As an instance to demonstrate the importance of shift-invariance property, Fig.1 shows two examples of 1-D shifted step responses. It is noted in (a) that magnitude of D8 DWT varies significantly across time-scale domain. The smoothly-varied response in (b) indicates the near shift-invariance of Gabor CWT. Similarly, this is also the fundamental assumption for supporting QWT to resolve the correspondence problem.



(a) (b)

Fig.1. (a) Magnitude of 1-D shifted step response of D8 DWT at three successive scales (b) Magnitude of 1-D shifted step response of Gabor CWT at three successive scales

2.2 Hilbert quadruple with no DC response

Four real 2-D function f_η, $\eta \in \{1,2,3,4\}$ are called Hilbert quadruple if

$$\mathrm{I}[(f_\kappa)^q_{\mathrm{A}}] = f_\lambda;\quad \mathrm{J}[(f_\kappa)^q_{\mathrm{A}}] = f_\mu;\quad \mathrm{K}[(f_\kappa)^q_{\mathrm{A}}] = f_\nu \tag{1}$$

for some permutation of pairwise different $\kappa,\lambda,\mu,\nu \in \{1,2,3,4\}$, where

$$f^q_{\mathrm{A}} = f + if_{Hi_x} + jf_{Hi_x} + kf_{Hi_{xy}} \tag{2}$$

Operators $\mathrm{I}(\bullet),\mathrm{J}(\bullet),\mathrm{K}(\bullet)$ in (1) respectively extract the three imaginary parts of the quaternion analytic signal $(f_\kappa)^q_{\mathrm{A}}$, which are in turn related to quaternion units i,j,k ($i^2 = j^2 = k^2 = -1$ and $ij = k$). Subscripts Hi_x, Hi_y in (2) respectively represent the partial Hilbert transform along x-axis and y-axis, while Hi_{xy} denotes the total Hilbert transform in x-y coordinates. Due to the commutativity of convolution and Hilbert transform, we can first use Hilbert transform to build a quaternion wavelet through real-valued filters and then construct analytic signal with convolution operator:

$$\begin{aligned} f^q_{\mathrm{A}} &= \psi^q_{\mathrm{A}} \otimes f \\ &= (\psi + \psi_{Hi_x} + \psi_{Hi_y} + \psi_{Hi_{xy}}) \otimes f \end{aligned} \tag{3}$$

where ψ is a real-valued wavelet with shift-invariance property as analyzed in section 2.1. It is noted that we should use (3) to perform multiscale signal decomposition for extracting complete frequency information of the original signal f.

As mentioned above, we can use Hilbert transform to build a quaternion wavelet for multiscale analytic signal analysis from a real-valued wavelet. In this configuration, the complete 2-D analytic signal contains one part in the 2-D frequency domain indexed by (u,v): the one having spectrum on the upper right quadrant, that is $u \geq 0$ and $v \geq 0$.

In addition, no DC response is expected in the filtering output. Because of the substantial power in natural signals at low frequencies this DC sensitivity often introduces a positive bias in the real part of the response. This is the main reason that most of the authors select the Gabors with small bandwidth (usually less than one octave).

2.3 Short-length filters with good localization in space-frequency domain

With respect to the quaternion wavelets, it is generally accepted that the measurement of disparity should require only local support in space-frequency domain. Uncertainty principle defines fundamental lower bound for the joint localization of a signal in spatial and frequency domain. Quaternion Gabor filters are appropriate when one is interested in local spectral properties of a signal since they fulfill the uncertainty relation as an equality [1]. However, they usually require heavy computations especially for the calculation of the quaternion Fourier transform. The calculation of disparity estimation should be quick. As a substitute for the quaternion Gabors, the method in [6] formed a pixel-wise quaternion wavelet through imposing Hilbert transform on short-length biorthogonal wavelets. In the following experiments, this kind of quaternion wavelets is used to extract multiscale 2-D phase structures.

3 Cooperative stereo matching using quaternion wavelets and top-down segmentation

If the two views in a QWT image pair are related as follows

$$O_l(x,y) = O_r(x + d_1(x,y), y + d_2(x,y)), \tag{4}$$

where the quaternion output is formed in the polar coordinates as $O = \rho e^{i\phi} e^{k\psi} e^{j\theta}$ and subscripts l and r represent the left view and the right view, the 2-D disparity $\mathbf{d} = \begin{pmatrix} d_1(x,y) \\ d_2(x,y) \end{pmatrix}$ at each pixel except those near phase singularities can be estimated by the quaternion phase differences as in most QWT-based matching methods [1][2][5]:

$$d_1(x,y) = \frac{\phi_l(x,y) - \phi_r(x,y) + \pi(2n+k)}{\frac{1}{2}\left(\frac{\partial}{\partial x}\phi_l(x,y) + \frac{\partial}{\partial x}\phi_r(x,y)\right)}, \tag{5}$$

$$d_2(x,y)=\frac{\theta_l(x,y)-\theta_r(x,y)+m\pi}{\frac{1}{2}(\frac{\partial}{\partial y}\theta_l(x,y)+\frac{\partial}{\partial y}\theta_r(x,y))} \quad (6)$$

Depending on m, k is defined as

$$k=\begin{cases}0, & \text{if } m \text{ is even}\\ 1, & \text{if } m \text{ is odd}\end{cases}, \quad (7)$$

where m and n count the number of phase wrapping circles, and k is to keep the sign of $O_l(x,y)$ and $O_r(x,y)$ unchanged.

The above phase-difference disparity model computes disparity pixel by pixel, not utilizing the general prior model of the scenes such as piece-wise smoothness constraint and uniqueness constraint. Similar to the work in [6], we formulate the correspondence problem as a cost minimization process and propose an improved version by introducing developments in preserving disparity discontinuity, reducing ambiguity and improving accuracy. The details will be discussed in section 3.1-3.2. Still, we use multiscale phase unwrapping to enlarge the disparity search scope and accelerate the convergence speed.

3.1 Segmentation-based support

To obtain better performance, most recent matching methods introduce global optimization techniques into disparity calculation. Without the recovery of epipolar geometry in uncalibrated stereo, QWT-based matching methods perform direct 2D disparity search in image space. In this case, the computational complexity would increase abruptly if global optimization is exploited. On the other side, local winner-take-all disparity optimization accounts for the fast performance seen in realtime matchers but sensitive to the window size for correspondence operations.

In this paper, we take advantages of object segmentation that reflects global aspects of the images and fast disparity assignment in local segment-based supports. It is observed that segmentation captures perceptually important groupings within which smoothness assumption of disparity is acceptable and large disparity discontinuities often occur on the boundaries of homogeneous luminous segments. Based on this global model of disparity distribution, the stereo matching problem becomes assigning the corresponding disparity plane to each local segment.

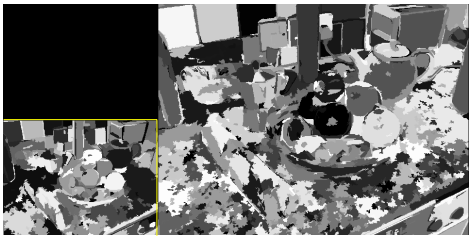


Fig.2. Top-down segmentation in two-level pyramid structure

In the proposed matching scheme, we execute top-down segmentation in the multiscale image pyramid. At each layer, winner-take-all disparity optimization is carried out based on the segment-based supports. As a result, disparity estimation is dependent on the contribution of the neighboring pixels in the same segment and consequently over-smoothness in discontinuity regions could be greatly alleviated between different segments. To get efficient segmentation at video rates, we exploit the idea of the work in [8]. Fig.2 shows the two-layer segmentation results for the uncalibrated image pair 'Fruit'.

3.2 Robust cost function

Given the candidate disparity $\mathbf{d}=(d_x,d_y)^T$ and the reference point $\mathbf{x}=(x,y)^T$ in the left view, the cost function in our approach is defined as:

$$\mathrm{Cost}(\mathbf{x},\mathbf{d})=\frac{\sum_{\Delta\mathbf{x}\subseteq(\Omega_{S\cap P})}[\rho_l(\mathbf{x}+\Delta\mathbf{x})\rho_r(\mathbf{x}+\Delta\mathbf{x}+\mathbf{d})\Delta\Phi(\mathbf{x},\Delta\mathbf{x},\mathbf{d})]}{\frac{1}{N}\sum_{-\mathbf{D}+\mathbf{d}\le\Delta\mathbf{x}'\le\mathbf{D}+\mathbf{d},\Delta\mathbf{x}'\neq\mathbf{0}}\Delta\Phi(\mathbf{x},\Delta\mathbf{x}',\mathbf{d}-\Delta\mathbf{x}')} \quad (6)$$

where

$$\begin{aligned}\Delta\Phi(\mathbf{x},\Delta\mathbf{x},\mathbf{d})&=\left|[\phi_l(\mathbf{x}+\Delta\mathbf{x})-\phi_r(\mathbf{x}+\Delta\mathbf{x}+\mathbf{d})]_{2\pi}\right|\\&+\left|[\theta_l(\mathbf{x}+\Delta\mathbf{x})-\theta_r(\mathbf{x}+\Delta\mathbf{x}+\mathbf{d})]_{\pi}\right|+\left|[\psi_l(\mathbf{x}+\Delta\mathbf{x})-\psi_r(\mathbf{x}+\Delta\mathbf{x}+\mathbf{d})]_{\pi/2}\right|\end{aligned} \quad (7)$$

Operator $[\bullet]_a$ in (7) is used to evaluate the principal value of $\bullet$ within the range of $[-\frac{a}{2},\frac{a}{2})$. In (6), $\Omega_{S\cap P}$ defines a support system for reliable cost aggregation. The influence of those pixels would be diminished within the aggregation support if they fall outside the image segment S that contains the central pixel under consideration. With regard to the slanted segment with notable disparity variations, we also exclude those outliers which fall outside the disparity plane P of the central pixel beyond a well-defined distance metric. This kind of outlier rejection operates in the multiscale matching structure except at the coarsest scale, where the initial disparity is not known.

Phase stability constraint is embedded in the above cost function to alleviate the singularity problem. The normalized amplitude ρ acts as weight factors in the cost evaluation, as a result the high-confidence matches in stable phase regions serve as heavyweight seeds for disparity propagation within the aggregation support.

Finally, we adopt the 'inhibition region' concept of cooperative stereo matching to enforce bidirectional uniqueness constraint on best disparity selection. If we only consider the minimization of the numerator cost term in (6), the optimal disparity is selected at the position where the quaternion phase structure in the right view is the most similar to the one under consideration in the left view. The inhibition region formed from the similar left sightlines is suppressed in minimizing this term. For cross validation, the denominator in (6) evaluates the costs in the inhibition region formed from similar right sightlines, where N counts the pixels number in this region and $\mathbf{D}$ defines the maximal 2-D disparity search range. As a result, minimization of the cost function in (6) implicitly imposes the bidirectional uniqueness on the disparity computation.

Given the cost function, we perform winner-take-all cost minimization process to estimate the optimal disparity map. The disparity at the coarse scale would serve as the initial disparity map of the next finer scale.

4 EXPERIMENTAL RESULTS

We use three image pairs to evaluate the matching performance of the proposed approach, including uncalibrated images 'Fruit', calibrated images 'Venus' with slanted surfaces as well as discontinuity regions and the calibrated aerial images 'pentagon' with illumination variations and noises. In comparison with the matching methods in [5] and [6], three horizontal disparity maps of 'Fruit' are shown in Fig.3. Subfigure (b) is computed by the proposed approach and by visual evaluation the discontinuity of disparity is better preserved than the result in [6], as compared in (e). Chan's method is suitable for very small shifts and loses its applicability to large and dense disparity estimation as shown in (d), where 2-D disparities are marked as vector arrows overlaid on the original image. In figure 4, we show the computed 1-D disparity maps of two calibrated image pairs to testify the robustness of the proposed approach. It is observed that the discontinuity is well preserved and disparity variations within the surfaces are captured. Compared with the disparity ground truth of 'Venus', there are 2.6 percent mismatches in non-occluded regions and 19.9 percent mismatches in discontinuity regions[1].

5 CONCLUSION

In this paper, the principles of quaternion wavelet construction are discussed in detail for achieving multiscale analysis of geometric image features. Then the quaternion wavelets are applied to propose a cooperative stereo matching algorithm using top-down segmentation-based disparity propagation. It aims to achieve smooth disparity map with details retained in discontinuity regions and slanted surfaces. As shown in above experiments, the matching results are improved in a comparison with the current QWT-based matching methods. In the future work, we would focus on global matching among segments to improve the work further. Due to the segments are much fewer than the pixels, computational complexity can be expected not too high.

6 ACKNOWLEDGEMENT

This work was supported by Shanghai Key Laboratory of Digital Media Processing and Transmission, National Natural Science Foundation of China (60502034), Shanghai Rising-Star Program (05QMX1435), Shanghai Postdoctoral Foundation (06R214138), Hi-Tech Research and Development Program of China (863: 2006AA01Z124), and the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (706022).

7 REFERENCES

[1] T. Bülow, "Hypercomplex spectral signal representations for the processing and analysis of images," Dessertation, Kiel: Institut für Informatik und Praktische Mathematik der Christian-Albrechts-Universität zu Kiel, 1999.

[2] E.B. Corrochano, "The theory and use of the quaternion wavelet transform," J. of Math. Imaging and Vision, vol.24, pp. 19-35, 2006.

[3] C. Philippe and D. Patrice, "Quaternionic wavelet transform for colour images," Proc. of the SPIE, vol. 6383, pp. 638301-1-638301-15, invited paper, 2006.

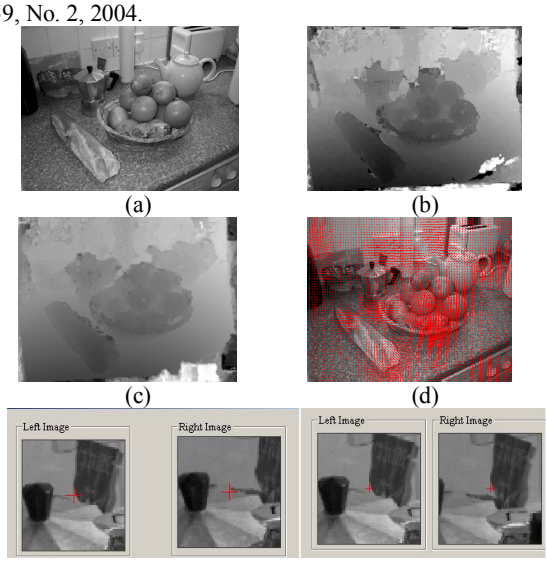
[4] C. Xie, M. Savvides and B.V.K. Vijaya Kumar, "Quaternion Correlation Filters for face Recognition in Wavelet domain," International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp.85-88, 2005.

[5] W.L. Chan, H. Choi and R.G. Baraniuk, "Coherent image processing using quaternion wavelets," Proc. of the SPIE, vol. 5914, pp. 59140Z-1-59140Z-10, invited paper, 2005.

[6] Y. Xu, J. Zhou and G.T. Zhai, "2D Phase-Based Matching in Uncalibrated Images," IEEE workshop on Signal Processing Systems, pp. 325-330, Athens, Greece, 2005.

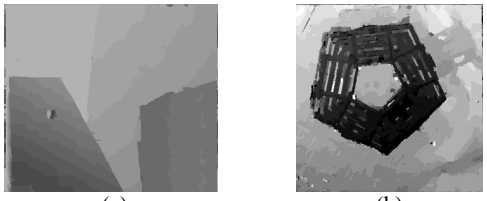
[7] N. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," Applied and Computational Harmonic, vol. 10, no. 3, pp. 234-253, 2001.

[8] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Graph-Based Image Segmentation," Int. J. of Computer Vision, vol. 59, No. 2, 2004.



(a) (b) (c) (d) (e) Two matches labeled by red crosses near the discontinuity respectively computed by the method in [6] (left) and the proposed method (right)

Fig.3 Related matching results in uncalibrated image pair 'Fruit' [2]



(a) (b)

Fig.4 Matching results by the proposed approach in calibrated images 'Venus' and 'Pentagon'

[1] Ground truth can be found in the open database: www.middlebury.edu/stereo

[2] The source code of Chan's method is open at www.dsp.ece.rice.edu/software