Stereo Video Disparity Estimation Using Multi-wavelets

[1]Pooneh Bagheri Zadeh and [2]Cristian V. Serdean
Department of Engineering, Faculty of Technology,
De Montfort University
Leicester, UK
E-mail: [[1]pbz, [2]cvs]@dmu.ac.uk

***Abstract*—Disparity estimation in stereo video processing is a crucial step in the generation of a 3D view of a scene. In this paper, a multi-wavelet based stereo correspondence matching technique for video is proposed. A multi-wavelet transform is first applied to a pair of stereo frames. Correspondence matching is initially performed at the coarsest level and relies on coarse-to-fine refinement in order to reduce the overall computational costs. Correspondence matching is carried out using a global error energy minimization technique to generate a disparity map for each of the four multiwavelet basebands of the stereo pair. Information in the resulting disparity maps is then combined using an interpolation operator to construct an initial disparity map. The information in the initial disparity map is then progressively propagated to higher resolution levels, on a coarse-to-fine basis, leading to a dense disparity map. Experimental results were generated using two sets of wide baseline convergent multi-view test videos: Breakdancers and Ballet. Results show that multi-wavelets can be a serious contender to scalar wavelets, producing smoother disparity maps with less mismatch errors compared to applying the same global error energy minimization algorithm in the wavelet domain.**

Keywords- Multi-wavelets, Correspondence matching, Disparity estimation, Stereo video.

I. INTRODUCTION

Recent years have seen significant advances in multimedia technologies with stereo video and 3D-TV equipment becoming a familiar consumer presence. In stereo video, three-dimensional perception is achieved by simultaneously providing two views of a scene captured by a stereoscopic camera to each corresponding eye, typically with the aid of a pair of active glasses. The brain will then process this stereo information and based on the disparities between the corresponding elements of the two scenes 'convert' it into a meaningful 3D internal representation. The key and the most complex operation in any stereoscopic to real 3D video system is disparity estimation, which needs to accurately find the correspondence points between the two stereo pairs and generate a disparity map for each frame pair. Using these disparity vectors in conjunction with the relevant camera parameters allows one to reconstruct a 3D model of a scene via conventional triangulation techniques.

Many stereo correspondence matching algorithms have been proposed over time, from feature based algorithms, to block-matching, pel-recursive, optical flow, and Bayesian-based approaches. Chien et al. [1] proposed a disparity estimation algorithm for mesh-based stereo images and video featuring a two-stage hybrid approach. In the first stage, an initial disparity map is generated using an iterative block matching algorithm. In the second stage, an iterative octagonal matching algorithm is employed to refine the disparity vectors. Another disparity estimation algorithm for stereo video based on epipolar geometry was reported by Lu et al. [2]. Their theoretical analysis and experimental results showed that their algorithm greatly reduce the search cost, while effectively tracking large and irregular disparities and being less sensitive to epipolar geometry estimation noise. Fan et al. [3] presented a disparity estimation algorithm for stereo video based on edge detection. This algorithm employs the characteristics of human visual system to reduce distortion around the edge regions. They report significant improvement in disparity estimation compared to other state of art techniques. Another disparity estimation technique for stereo video was proposed by Zhu et al. [4], which employed both spatio-temporal correlation and temporal variation of disparity field techniques. By using this technique, they achieved an important reduction in computational complexity compared to full search algorithms.

Over the past years much research has been done to improve the performance of correspondence matching techniques, as well as reducing the computational cost of the search for the best match. Multi-resolution based stereo matching algorithms have received much attentions due to the hierarchical and scale-space localization properties of the wavelets [5],[6]. Correspondence matching can be performed hierarchically, leading to lower computational costs. Yongdong and Guiling [7] proposed a hierarchical multi-resolution based block matching technique for disparity estimation in stereo video. They reported significant improvement in the smoothness of disparity field as well as a reduction in the computational load. Sarkar and Bansal [6] presented a multi-resolution based correspondence matching technique using a mutual information algorithm. They showed that such technique can produce significantly more accurate matching results compared to conventional correlation based algorithms at lower computational costs.

Multi-wavelets offer a number of desirable properties compared to scalar wavelets such as their ability to possess orthogonality, symmetry and high orders of approximation all at once [8]. These properties could increase the accuracy of correspondence matching techniques while still exploiting

 ISBN: 978-1-61208-193-9

their hierarchical nature in order to reduce the overall complexity of the correspondence matching algorithms via coarse-to-fine refinement. Bhatti and Nahavandi [9] proposed a multiwavelet based stereo correspondence matching algorithm which makes use of the wavelet transform modulus maxima to generate a disparity map at the coarsest level. This is then followed by a coarse-to-fine strategy to refine the disparity map up to the finest level. Bagheri Zadeh and Serdean [10] provided an evaluation on different types and families of multiwavelets in stereo correspondence matching. They developed an algorithm based on normalized cross correlation. To generate a dense disparity map from the four basebands, a shuffling technique was used in case of balanced multiwavelets and a Fuzzy algorithm was employed in the case of unbalanced multiwavelets. Results showed that the unbalanced multiwavelets produced a smoother disparity map with less mismatch errors compared to balanced multiwavelets.

This paper presents a multi-wavelet based stereo matching algorithm for video which employs a global error energy minimization technique. A multi-wavelet transform is first applied to the input stereo pair to decompose them into a number of subbands. The global error energy minimization algorithm is then employed to generate a disparity map using the coarse subbands. A median operator is then used to combine the disparity maps and generate an initial disparity map. The estimated disparity map is then refined at higher resolution levels, taking advantage of the hierarchical, multi-resolution nature of the multiwavelets to efficiently generate a more accurate final disparity map.

The paper is organized as it follows. Section II presents a brief review of the multi-wavelet transform. The proposed stereo matching technique is discussed in Section III. Experimental results are presented in Section IV while Section V is dedicated to the conclusions.

II. MULTIWAVELET TRANSFORM

Multi-wavelet transforms are similar to scalar wavelet transforms with some key differences. Classical wavelet theory is based on the refinement equations as given below:

$$\begin{aligned} \phi(t) &= \sum_{k=-\infty}^{k=\infty} h_k \phi(m\,t-k) \\ \psi(t) &= \sum_{k=-\infty}^{k=\infty} g_k \psi(m\;t-k) \end{aligned} \tag{1}$$

where $\phi(t)$ is a scaling function, $\psi(t)$ is a wavelet function, h_k and g_k are scalar filters and m represents the band number.

In contrast to wavelet transforms, multi-wavelets have two or more scaling and wavelet functions. Scalar wavelets have multiplicity $r=1$, while multi-wavelets support $r \geq 2$. To date, most multiwavelets have a multiplicity factor of $r=2$.

The set of scaling and wavelet functions of a multi-wavelet in vector notation can be defined as:

$$\begin{aligned} \Phi(t) &\equiv [\phi_1(t) \quad \phi_2(t) \quad \phi_3(t) \quad ... \quad \phi_r(t)]^T \\ \Psi(t) &\equiv [\psi_1(t) \quad \psi_2(t) \quad \psi_3(t) \quad ... \quad \psi_r(t)]^T \end{aligned} \tag{2}$$

where $\Phi(t)$ and $\Psi(t)$ represent the multi-scaling and respectively multi-wavelet functions, with r scaling- and wavelet functions. A multi-wavelet with two scaling and wavelet functions can be denoted as [11]:

$$\begin{aligned} \Phi(t) &= \sqrt{2} \sum_{k=-\infty}^{k=\infty} H_k \Phi(m\,t-k) \\ \Psi(t) &= \sqrt{2} \sum_{k=-\infty}^{k=\infty} G_k \Psi(m\,t-k) \end{aligned} \tag{3}$$

where H_k and G_k are $r \times r$ matrix filters and m is the subband number.

Unlike scalar wavelets, multi-wavelets can offer symmetry, orthogonality and approximation orders higher than 1 simultaneously. Similar to wavelet transforms, multi-wavelets can be implemented using Mallat's filter bank theory [5]. A 2D multi-wavelet transform with multiplicity two will produce sixteen subbands: four basebands and twelve high frequency subbands. A visual comparison of the resulting subbands of a 2D wavelet (Antonini 9/7) and respectively 2D multi-wavelet (bat01) is shown in Figure 1.

III. MULTI-WAVELET IN STEREO VIDEO CORRESPONDENCE MATCHING

A block diagram of the multi-wavelet based stereo correspondence matching technique for stereo video based on a global error energy minimization algorithm is shown in Figure 2. A stereoscopic video needs to be input to the system. For the purpose of this paper, two camera views from the multi-view sequences, Breakdancers and Ballet (generated by Microsoft laboratories using eight synchronized PtGrey color cameras) are chosen [12]. As these datasets were captured using convergent cameras, each frame pair needs to be rectified to suppress the vertical displacement. The epipolar rectification algorithm proposed by Fusiello and Irsara [13] has been used in this work to rectify each fame pair of the video input. A multi-wavelet transform is then applied to each rectified frame pair to decompose them into multi-wavelet subbands. The search for the best correspondence points starts at the coarsest level. The corresponding basebands in the two frames are passed to a regional based stereo matching block. The matching algorithm uses a global energy minimization technique [14] to generate a disparity map between the two input subbands. This global error energy minimization technique is briefly described in Section III.A. The output of the matching process is four disparity maps. These maps are then combined using a median operator to generate an initial disparity map. As the initial disparity map is estimated at the lowest resolution, the information needs to be progressively passed on to higher resolution levels. For this refinement

 ISBN: 978-1-61208-193-9

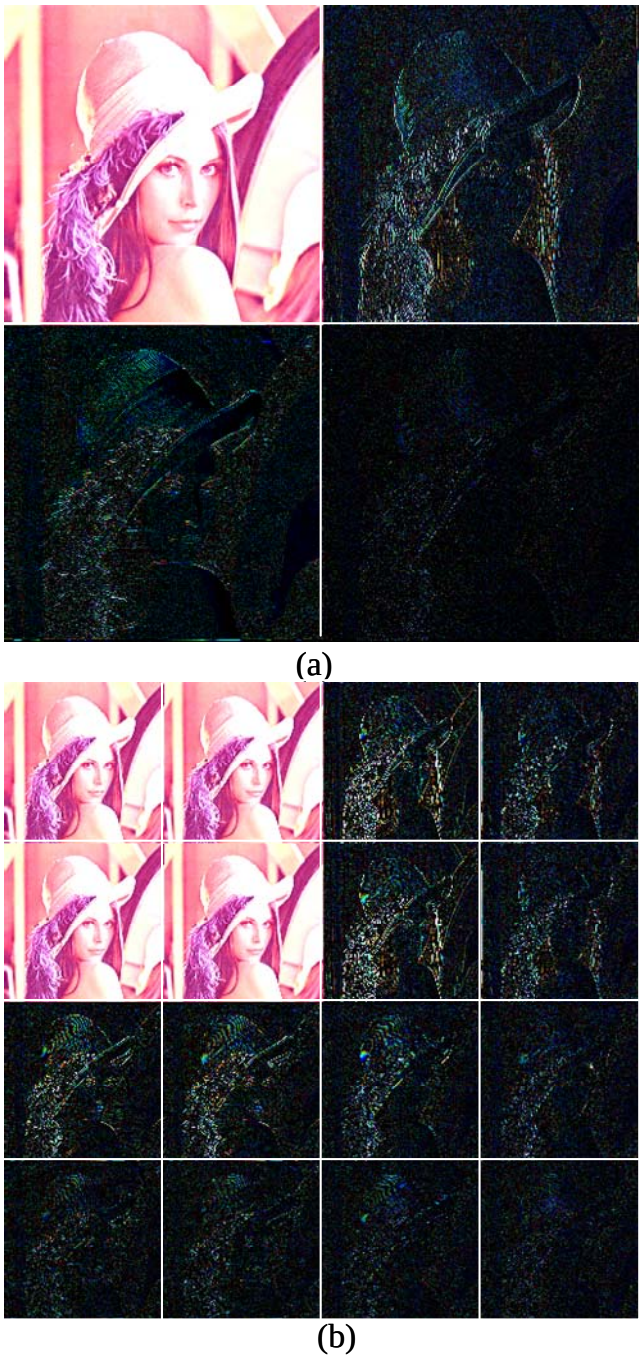


Figure 1. Single level decomposition of Lena test image (a) Antonini 9/7 wavelet transform (b) bat01 multi-wavelet transform.

process, the algorithm presented in [6] is used to propagate information in the coarsest level to the higher resolutions. Finally a median filter is applied to the last processed disparity map to further smooth the final disparity map.

A. *Global Error Energy Minimization technique*

The Global Error Energy Minimization (GEEM) technique [14] calculates a disparity vector for each pixel. It searches for the best match for each pixel in the correspondence

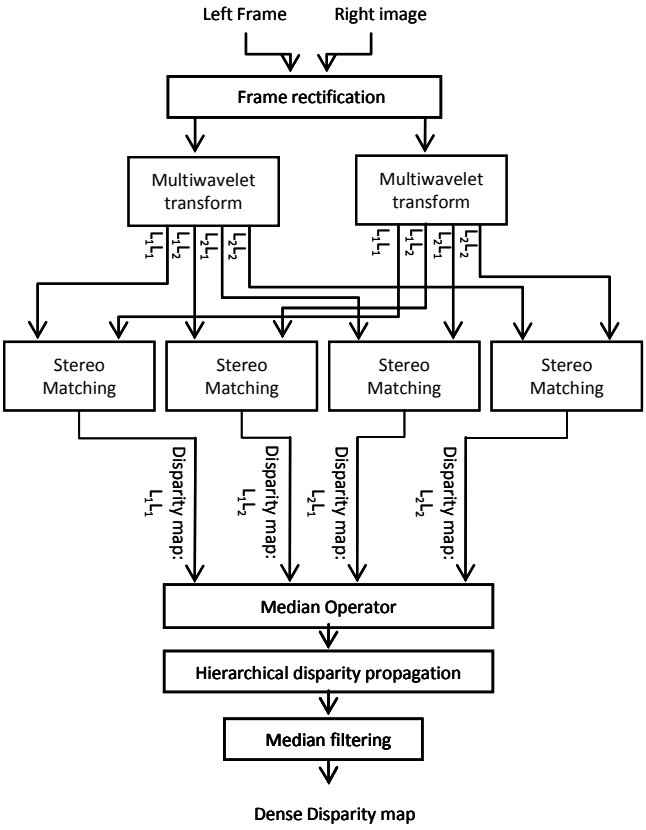


Figure 2. Block diagram of multi-wavelet based stereo matching technique using the global energy minimization algorithm.

search area of the other image using error minimization criterion. For RGB images, the error energy criterion can be defined as:

$$Er_{en}(i,j,w_x,w_y) = \frac{1}{3}\sum_{k=1}^{3}\left(I_1(i+w_x,j+w_y,k)-I_2(i,j,k)\right)^2$$
$$-d_x \le w_x \le d_x \quad and \quad -d_y \le w_y \le d_y$$
$$i = 1, \ldots m \quad and \quad j = 1, \cdots n \qquad (4)$$

where I_1 and I_2 are the two input frames, $Er_{en}(i,j,w_x,w_y)$ is the difference energy of the pixel $I_2(i,j)$ and pixel $I_1(i+w_x,j+w_y)$, d_x is the maximum displacement around the pixel in the x direction, d_y is maximum displacement around the pixel in the y direction and m and n are the image size.

In order the GEEM algorithm to determine the disparity vector for each pixel in the current view, it first calculates Er_{en} of each pixel with all the pixels in its search area in the corresponding frame. For every disparity vector (w_x,w_y) in the disparity search area, error energy is calculated using Equation 4 and placed into a matrix. Each of the resulting error energy matrices is first filtered using an average filter

 ISBN: 978-1-61208-193-9

Camera 4, Frame25 Camera 5, Frame 25 Camera 4, Frame99 Camera 5, Frame 99

(a) 'Ballet' Multi-view Sequence

Camera 4, Frame 51 Camera 5, Frame 51 Camera 4, Frame99 Camera 5, Frame 99

(b) 'Breakdancer' Multi-view Sequence

Figure 3. Two views of Multi-view test sequences, (a) 'Ballet' and (b) 'Breakdancers.

to decrease the number of incorrect matches [15]. The disparity index of each pixel is then determined by finding the disparity index of the matrix which contains the minimum error energy for that pixel. In order to increase the reliability of the disparity vectors around the object boundaries, which is the result of object occlusion in images, the generated disparity map undergoes a thresholding procedure as it follows:

$$\tilde{d}(i,j) = \begin{cases} d(i,j) & Er_{en}(i,j) \le \alpha \times Mean(Er_{en}) \\ 0 & Er_{en}(i,j) > \alpha \times Mean(Er_{en}) \end{cases} \quad (5)$$

where $\tilde{d}(i,j)$ is the processed disparity map, $d(i,j)$ is the disparity map, α is a tolerance reliability factor, $Er_{en}(i,j)$ is the minimum error energy of the pixel (i,j) calculated and selected in the previous stage. Finally a median filter is applied to the processed disparity map, $\tilde{d}(i,j)$ to further smooth the final disparity map.

IV. Simulation Results

In order to evaluate the performance of the proposed multi-wavelet technique compared to a similar wavelet based stereo matching technique which employs the same global energy minimization algorithm, both methods are benchmarked using two multi-view sequences, Breakdancers and Ballet [12]. Figure 3 shows frames 25 and 99 of the camera 4 and camera 5 views of Ballet sequence and respectively frames 51 and 99 of the two views (camera 4 and camera 5 views) of Breakdancers video. The resulting disparity maps for the two methods using the GHM unbalanced multi-wavelet and respectively the Antonini 9/7 scalar wavelet for the Ballet sequence (frame 25 and 99 from camera views of 4 and 5) and Breakdancers sequence (frame 51 and 99 from camera views of 4 and 5) are illustrated in Figures 4(a) and 4(b). In these figures areas with intensity 0 represent unreliable disparities. From Figure 4, it is obvious that the disparity map produced by the multi-wavelet based algorithm is more accurate and smoother than that of the wavelet based technique. The different spectral content of the multiwavelet subbands and the greater subband structure flexibility afforded by multi-wavelets enable the global energy minimization algorithm to generate more reliable matches from the multi-wavelet decomposition than from the scalar wavelet decomposition.

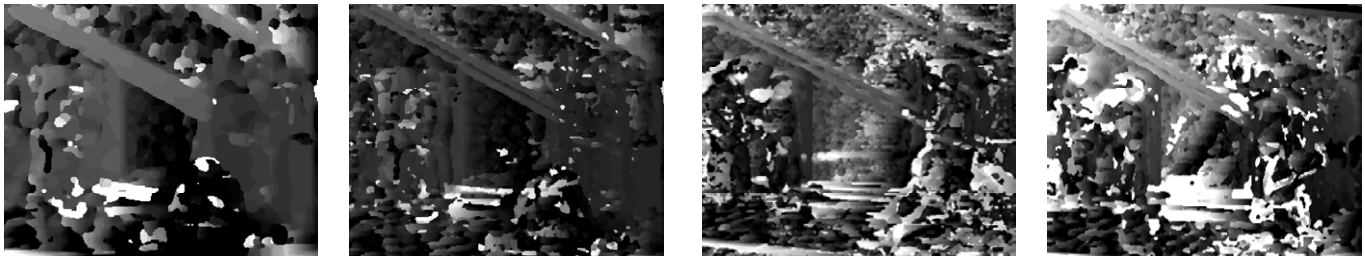
V. Conclusion

This paper presented a multi-wavelet based stereo correspondence matching technique for video using a global error energy minimization algorithm. A multi-wavelet transform with multiplicity of two decomposes the input rectified frame pairs into four baseband and twelve high frequency subbands. The resulting four basebands of the two views were then employed to generate four disparity maps using the global error energy minimization algorithm. The resulting four disparity maps were then combined using a median operator to generate the initial disparity map, which was then refined by hierarchically propagating it to the finer levels. Results show that multi-wavelets can be a serious

 ISBN: 978-1-61208-193-9

Multi-wavelet, Frame 25 Wavelet, Frame 25 Multi-wavelet, Frame 99 Wavelet, Frame 99

(a) 'Ballet' Multi-view Sequence



Multi-wavelet, Frame 51 Wavelet, Frame 51 Multi-wavelet, Frame99 Wavelet, Frame 99

(b) 'Breakdancer' Multi-view Sequence

Figure 4. Disparity maps generated using wavelet based algorithm and the multi-wavelet based algorithm for (a) 'Ballet' multi-view sequence and (b) 'Breakdancers' multiview sequence.

contender to scalar wavelets, producing smoother disparity maps with less mismatch errors compared to applying the same global error energy minimization algorithm in the wavelet domain.

ACKNOWLEDGMENT

This work was supported by EPSRC under Grant number: EP/G029423/1.

REFERENCES

[1] S. Chien, S. Yu, L. Ding, Y. Huang, and L. Chen, "Fast disparity estimation algorithm for mesh-based stereo image/video compression with two-stage hybrid approach," Proceedings of SPIE, Vol. 5150, pp. 1521-1530, 2003.

[2] J. Lu, H. Cai, J. G. Lou and J. Li, "An epipolar geometry-based fast disparity estimation algorithm for multiview image and video coding," IEEE Transaction on Circuits System for Video Technology, vol. 17, no. 6, pp.737–750, June 2007.

[3] J. Fan, F. Liu, W. Bao and H. Xia, "Disparity Estimation Algorithm for Stereo Video Coding Based on Edge Detecetion," International Conference on Wireless Communications & Signal Processing, pp. 1-5 , November 2009.

[4] W. Zhu, X. Tian, F. Zhou and Y. Chen, "Fast Disparity Estimation Using Spatio-temporal Correlation of Disparity Field for Multiview Video Coding," IEEE Transactions on Consumer Electronics, vol. 56, no. 2, pp. 957-964, 2010.

[5] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, 1999.

[6] I. Sarkar, M. Bansal, "A wavelet-based multiresolution approch to solve the stereo correspondence problem using mutual information," IEEE Transaction on system, man, and cybernetics, vol. 37, pp. 1009-1014, Auguest 2007.

[7] Z. Yongdong and L. Guiling, "The research of disparity estimation algorithms in stereo video coding," Journal of Electronic Measurement and Instrumentation, January 2002.

[8] V. Strela and A.T. Walden, "Signal and image denoising via wavelet thresholding: orthogonal and biorthogonal, scalar and multiple wavelet transforms," In Nonlinear and Nonstationary Signal Processing, pp. 124-157, 1998.

[9] A. Bhatti and S. Nahvandi, "Depth estimation using multi-wavelet analysis based stereo vision approach," International Journal of Wavelets, Multiresolution and Information Processing, vol. 6, pp. 481-497, 2008.

[10] P. Bagheri Zadeh and C. Serdean, "An Evaluation of Multiwavelet Families For Stereo Correspondence Matching", The sixth International Conference on Digital Telecommunications (ICDT2011), Budapest, Hungary, pp. 41-45, June 2010.

[11] V. Strela, "Multiwavelets: threory and applications," PhD thesis, MIT, 1996.

[12] C.L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *ACM SIGGRAPH and ACM Trans. on Graphics*, Los Angeles, CA, pp. 600-608, Aug. 2004.

[13] A. Fusiello and L. Irsara, "Quasi-euclidean Uncalibrated Epipolar Rectification," International Conference on Pattern Recognition (ICPR), 2008, Tampare, Finland, 2008.

[14] B. B. Alagoz, "Obtaning depth maps from colour images by region based stereo matching algorithms," OncuBilim Algorithm and System Labs, vol. 08, Art.No:04, 2008.

 ISBN: 978-1-61208-193-9