A HIERARCHICAL MULTIWAVELET BASED STEREO CORRESPONDENCE MATCHING TECHNIQUE

[1] *Pooneh Bagheri Zadeh and* [2]*Cristian V. Serdean*

Department of Engineering, De Montfort University
The Gateway, Leicester, LE1 9BH, United Kingdom
phone: +44 (0)116 257 7691, fax: +44 (0)116 257 7692, emails: [1]pbz@dmu.ac.uk, [2]cvs@dmu.ac.uk,
web: http://www.dmu.ac.uk/

ABSTRACT

This paper presents a hierarchical stereo correspondence matching technique based on multiwavelet transforms. A global error energy minimization technique is employed to generate a disparity map for each of the four multiwavelet approximation subband pairs. The information in the four disparity maps is then combined using a Fuzzy algorithm to generate a single disparity map. This initial disparity map is estimated at the lowest resolution and needs to be progressively passed on to higher resolution levels. Hence, the search at higher resolution levels is significantly reduced, thereby reducing the computational cost of the overall process and improving the reliability of the final disparity map. Results show that the proposed technique produces a smoother disparity map with less mismatch errors compared to applying the same method in both spatial and wavelet domains. The proposed algorithm fares very well when compared to other state of art techniques from the Middlebury database.

1. INTRODUCTION

Stereo correspondence aims to find the closest possible match between the corresponding points of two images captured simultaneously by two cameras placed at slightly different spatial locations. The cameras are usually aligned in such a way that each scan line of the rectified images corresponds to the same line in the other image, hence searching for the best correspondence match is done horizontally. A disparity map generated from the correspondence matching process, along with the stereo camera parameters are then used to calculate the depth map and produce a 3D view of the scene. However, a number of problems such as occlusion, ambiguity, illumination variation and radial distortion complicate the search for the best corresponding points between the two views [1].

Multiresolution analysis has played a significant role in stereo correspondence matching and 3D reconstruction, which has led to promising results especially in terms of creating a robust and dense disparity map [2-5]. This is due to the hierarchical and scale-space localization properties of the wavelets [4, 6]. This allows for correspondence matching to be performed on a coarse-to-fine basis, resulting in decreased computational costs. Jiang et al. proposed a wavelet based stereo image pair coding algorithm [2]. A wavelet transform decomposes the images into low and high frequency subbands and the disparity map is estimated using both the approximation and edge information. This is followed by a disparity compensation and subspace projection technique to improve the disparity map estimation. Caspary and Zeevi [3] proposed a wavelet based stereo matching technique which employs a differential operator in the wavelet domain to iteratively minimize a defined cost function. Sarkar and Bansal [4] introduced a multiresolution based correspondence matching technique using a mutual information algorithm. They showed that the multiresolution technique produces significantly more accurate matching results compared to correlation based algorithms at much lower computational cost. Li et al [5] present a stereo correspondence matching technique based on the 2D monogenic wavelet transform, which pairs the polyharmonic B-spline wavelet basis with its complex Riesz counterparts to specify a multiresolution monogenic signal analysis. They reported promising results compared to the state of art techniques.

Research has shown that unlike scalar wavelets, multiwavelets can possess orthogonality (preserving length), symmetry (good performance at the boundaries via linear-phase), and a high approximation order at the same time [7], which could potentially increase the accuracy of correspondence matching techniques.

Bhatti and Nahavandi [8] were amongst the very few to propose a multiwavelet based stereo correspondence matching algorithm. They use 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 [9] proposed another multiwavelet based stereo correspondence matching technique to compare balanced multiwavelets versus unbalanced multiwavelets. Overall, in spite of their highly desirable properties compared to scalar wavelets, multiwavelets have been still relatively little used in stereo correspondence matching algorithms so far.

In this paper, a novel multiwavelet based stereo correspondence matching algorithm using a global error energy minimization technique is presented. A multiwavelet is first applied to the input stereo images to decompose them into a number of subbands. A global error energy minimization algorithm is then employed to generate a disparity map

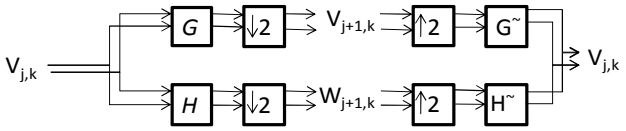


Figure 1 – Analysis / synthesis stage of one level multiwavelet transform.

for each of the four approximation subbands. A Fuzzy algorithm is used to combine the four disparity maps and generate an initial disparity map. The initial estimated disparity map is then refined at higher resolution levels to form the final disparity map.

The paper is organized as it follows. Section 2 introduces a brief review of the multiwavelet transform. The proposed stereo matching technique is discussed in Section 3. Experimental results are presented in Section 4 and the conclusions are outlined in Section 5.

2. MULTIWAVELET TRANSFORM

In many respects, multiwavelet transforms are very similar to scalar wavelet transforms. Classical wavelet theory is based on the following refinement equations:

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

where $\phi(t)$ is a scaling function, $\psi(t)$ is a wavelet function, h_k and g_k are scalar filters, m represents the subband number and k is the shifting parameter. In contrast to wavelet transforms, multiwavelets have two or more scaling and wavelet functions. The set of scaling and wavelet functions of a multiwavelet in vector notation can be defined as:

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

where $\Phi(t)$ and $\Psi(t)$ are the multi-scaling and respectively multiwavelet functions, with r scaling- and wavelet functions. In the case of scalar wavelets their multiplicity order is $r = 1$, while multiwavelets support $r \geq 2$. To date, most multiwavelets have $r = 2$. A multiwavelet with two scaling and wavelet functions can be defined as [10]:

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

where H_k and G_k are 2x2 ($r \times r$) matrix filters and m is the subband number [7]. Similar to wavelet transforms,

L_1L_1	L_1L_2	L_1H_1	L_1H_2
L_2L_1	L_2L_2	L_2H_1	L_2H_2
H_1L_1	H_1L_2	H_1H_1	H_1H_2
H_2L_1	H_2L_2	H_2H_1	H_2H_2

Figure 2 – One level of 2D Multiwavelet decomposition.

multiwavelets can be implemented using Mallat's filter bank theory [6]. Figure 1 shows one level of analysis/synthesis for a 1D multiwavelet transform, where blocks G and H are low- and high-pass analysis filters and $G^{\sim}$ and $H^{\sim}$ are low- and high-pass synthesis filters. Due to its separability property, a 2D multiwavelet transform can be implemented via two 1D transforms. Therefore, for one level of decomposition, a 2D multiwavelet with multiplicity of 2 generates sixteen subbands, as shown in Figure 2. In Figure 2, L_xL_y represent the approximation subbands, while L_xH_y, H_xL_y and H_xH_y are the detail subbands, with $x = \overline{1,2}$ and $y = \overline{1,2}$. A visual comparison of the resulting subbands for a 2D wavelet and respectively multiwavelet, is shown in Figures 3(a) and 3(b). As it can be seen from Figure 3, the multiwavelet transform generates four subbands instead of each subband created by the wavelet transform, and these four subbands carry different spectral content of the input image due to multiwavelet's filters properties.

The major advantage of multiwavelets over scalar wavelets is their ability to possess symmetry, orthogonality and higher order of approximation simultaneously, which is impossible for scalar wavelets. Furthermore, the multichannel structure of the multiwavelet transform is a closer approximation of the human visual system than what wavelets offer. This multichannel structure has the potential to increase the accuracy of the disparity calculation and reduce the number of erroneous matches in disparity maps. Further information about the generation of multiwavelets, their properties and their applications can be found in [7, 10].

3. MULTIWAVELET BASED STEREO CORRESPONDENCE MATCHING TECHNIQUE

Figure 4 shows a block diagram of the proposed multiwavelet based stereo matching technique. A pair of rectified stereo images is input to the system. A multiwavelet transform is then applied to the input stereo images to decompose them into a number of subbands. The pre-filter employed in this paper is a repeat row type. The information in the approximation subbands is less sensitive to the shift variability of the multiwavelets. The four approximation subbands are first used to generate an initial disparity map. The corresponding approximation subband pairs in the two images are passed to a regional based stereo matching block. The matching algorithm uses a global error energy minimization technique [11] to generate a disparity map between the two input subbands. This global energy minimization technique

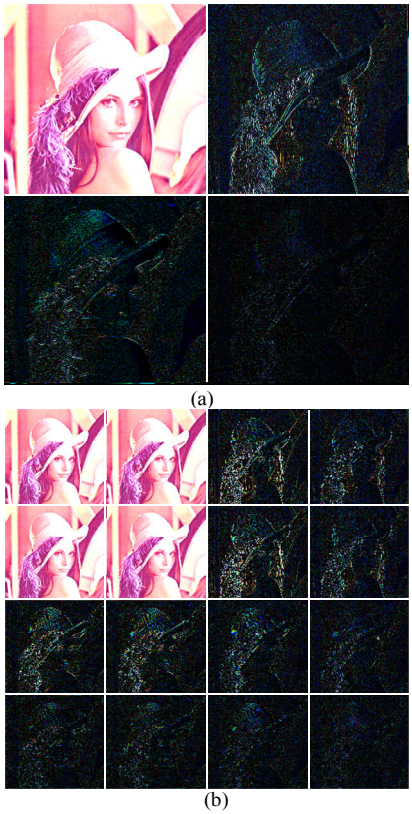


Figure 3 – Single level decomposition of Lena test image (a) Antonini 9/7 wavelet transform, (b) balanced bat01 multiwavelet transform.

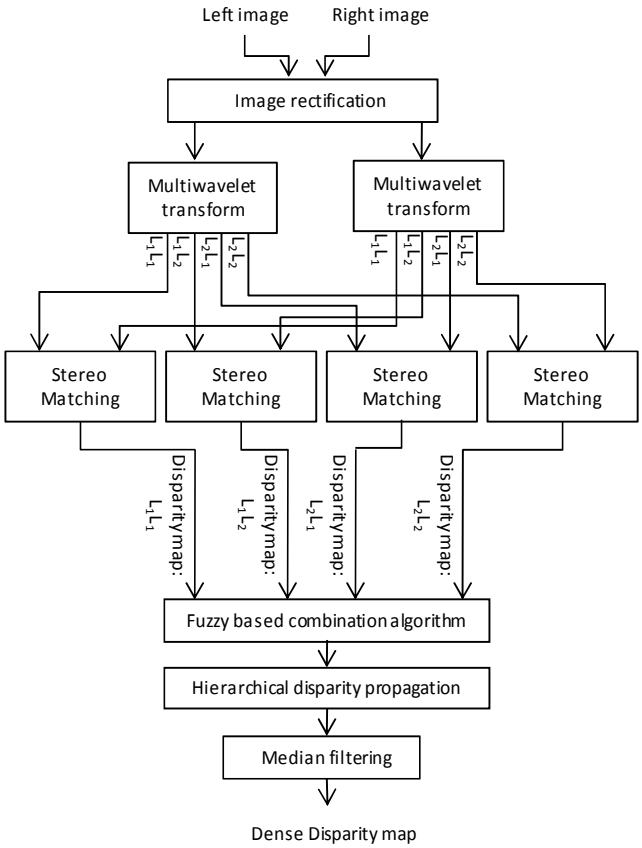


Figure 4 – Block diagram of the multiwavelet based stereo matching technique using the global error energy minimization algorithm.

is briefly described in sub-section 3.1. The matching process outputs four disparity maps. These maps are then combined using a Fuzzy algorithm to generate a dense initial disparity map, which reduces the number of erroneous matches. This Fuzzy algorithm combines the same disparity values $d_{L_1L_1}, d_{L_1L_2}, d_{L_2L_1}$ and $d_{L_2L_2}$ generated from L_1L_1, L_1L_2, L_2L_1 and L_2L_2 basebands respectively, with the view of giving a higher weight to the disparity values in $d_{L_1L_1}$. The disparity values from the other three disparity maps $d_{L_1L_2}, d_{L_2L_1}, d_{L_2L_2}$ are used to refine the initial disparity values. If the difference between the four disparity values is less than a threshold value (in this paper equal to 1), the mean value of the four disparities is put in the initial disparity map. In the other cases the initial disparity value for each pixel is calculated using the following empirical formula:

$$d_{\text{int}ial} = \alpha \times d_{L_1L_1} + \frac{d_{L_1L_2} + d_{L_2L_1} + d_{L_2L_2}}{3} \qquad (4)$$

where α is a weighting factor. If the difference between the value of each of the $d_{L_1L_2}, d_{L_2L_1}, d_{L_2L_2}$ disparities and the $d_{L_1L_1}$ disparity is larger than a threshold value (in this paper equal to 4), they are discarded in the above formula.
This initial estimated disparity map is generated at the lowest resolution level and it needs to be refined by progressively passing it on to higher resolution levels. Using the coarse-to-fine refinement principle, the search at high resolution levels can be significantly reduced, thereby reducing the computational cost of the overall process. The refinement process is detailed in sub-section 3.2. Finally a median filter is applied to the dense disparity map obtained as a result of hierarchical disparity propagation, which leads to a smoother, final disparity map.

3.1 Global Error Energy Minimization technique

The Global Error Energy Minimization (GEEM) technique [11] calculates a disparity vector for each pixel. It searches for the best match for each pixel in the correspondence search area of the other image using an error minimization criterion. For RGB images, the error energy criterion can be defined as:

$$Er_{en}(i,j,w_x,w_y) = \frac{1}{3}\sum_{k=1}^{3}(I_1(i+w_x,j+w_y,k)-I_2(i,j,k))^2$$
$$\begin{cases} -d_x \le w_x \le d_x \quad and \quad -d_y \le w_y \le d_y \\ i = 1,\ldots,m \quad and \quad j = 1,\ldots,n \end{cases} \qquad (5)$$

where I_1 and I_2 are the two input images, $Er_{en}(i,j,w_x,w_y)$ is the energy difference 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 x direction, d_y is the maximum displacement around the pixel in y direction, m and n are the image size and k represents the three components of an RGB image.

In order to determine the disparity vector for each pixel in the current view, the GEEM algorithm first calculates Er_{en} of each pixel with all the pixels from its search area in the correspondence image. For every disparity vector (w_x, w_y) in the disparity search area, the energy of the error is calculated using equation (5) and placed into a matrix. Each of the resulting energy error matrices is first filtered using an average filter to decrease the number of incorrect matches [12]. 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:

$$\tilde{d}(i,j) = \begin{cases} d(i,j) & Er_{en}(i,j) \le \alpha \times Mean(Er_{en}) \\ 0 & Er_{en}(i,j) > \alpha \times Mean(Er_{en}) \end{cases} \quad (6)$$

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

3.2 Hierarchical disparity propagation

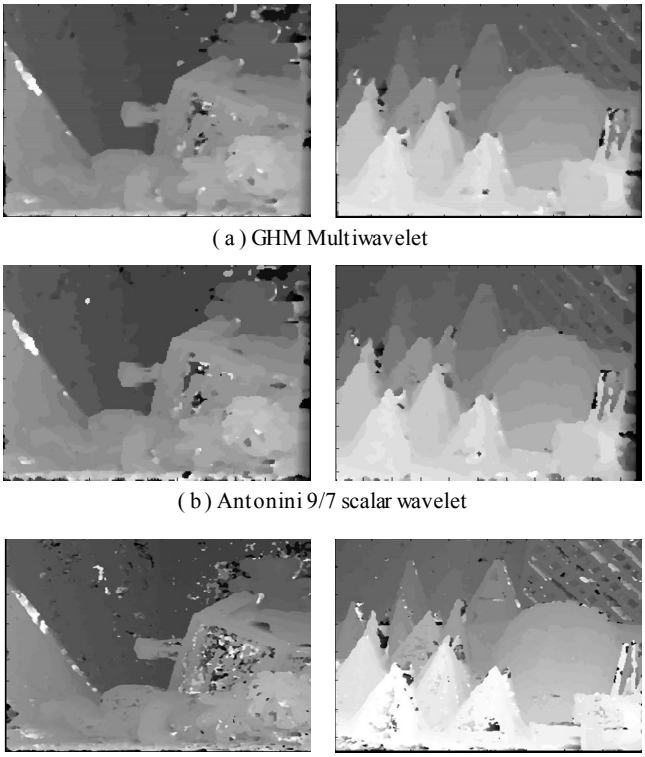
The information in the initial disparity map $\tilde{d}(i,j)$, generated at the coarsest level, needs to be propagated to the higher resolutions. Based on the wavelet theory, one point (x,y) of a coarse subband in the decomposition level $i+1$ corresponds to four points $(2x,2y)$, $(2x+1,2y)$, $(2x,2y+1)$ and $(2x+1,2y+1)$ of its finer subband at the decomposition level i. If (x,y) in the left image corresponds to (x',y') in the right image at level $i+1$, $(2x,2y)$ corresponds to one of the four points $(2x',2y')$, $(2x'+1,2y')$, $(2x',2y'+1)$ and $(2x'+1,2y'+1)$ from level i. Hence, the disparity in level $i+1$ can be propagated to the next finer level i by:

$$D_i(2x,2y) = 2D_{i+1}(x,y) + \Delta d \quad (7)$$

where Δd is one of $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$, which minimizes the error of the matching metric. Disparities at the remaining points are interpolated from $D_i(2x,2y)$. A similar scheme has also been employed by Sarkar and Bansal [4] in their paper.

4. SIMULATION RESULTS

The performance of the proposed algorithm has been assessed against the 'Cones', 'Tsukuba', 'Teddy' and 'Venus' stereo test images from the Middlebury stereo database [13]. In order to give a visual comparison, the performance of the proposed multiwavelet based GEEM algorithm is compared against similar GEEM algorithms operating in the spatial domain and respectively in the wavelet domain. The experimental results were generated using the unbalanced GHM multiwavelet and the Antonini 9/7 scalar wavelet. The resulting disparity maps obtained using the proposed multiwavelet based algorithm, the wavelet based algorithm and respectively the GEEM technique applied to the original stereo views for the 'Teddy' and 'Cones' stereo pairs, are illustrated in Figures 5(a), 5(b) and 5(c) respectively. In these figures areas with intensity zero represent occluded and unreliable disparities. As Figure 5 shows, the proposed multiwavelet based algorithm produces more accurate and smoother disparity maps compared to both wavelet and spatial domain GEEM based algorithms. This can be explained by the multichannel structure of the multiwavelet transform, where the four resulting subbands carry different spectral content of the input images, which in turn enables the global error energy minimization algorithm to generate more reliable matches than in the other two less 'adaptive' cases.



(a) GHM Multiwavelet

(b) Antonini 9/7 scalar wavelet

(c) spatial domain

Figure 5 – Disparity maps for stereo test images 'Teddy' (left) and 'Cones' (right) using: a) the proposed multiwavelet-based GEEM, b) the wavelet-based GEEM and c) GEEM in spatial domain.

In order to give an objective quality comparison, the proposed algorithm is also evaluated against some well known techniques from the Middlebury database [13]. The results are presented in Table 1. The chosen algorithms used for comparison are: AdaptingBP [14] (ranked second in the Middlebury database), DoubleBP [15] (ranked fourth in the Middlebury database), Graph Cut [16] and DP [17]. Table 1 shows the percentage of "bad pixels" at which the disparity error is bigger than 1. For each pair of images, the results in non-occluded regions (nonoc.), all regions (all) and depth discontinuity regions (disc.) are presented. From Table 1, it

Algorithm	'Tsukuba'		
	Nonoc.	All	Disc.
Proposed method	0.89	1.39	5.9
AdaptingBP	1.11	1.37	5.79
Double BP	0.88	1.29	4.76
Graph cut	1.27	1.99	6.48
DP	4.12	5.04	12
	'Venus'		
Proposed method	2.59	2.61	2.02
AdaptingBP	0.1	0.21	1.44
Double BP	0.13	0.45	1.87
Graph cut	2.79	3.13	3.6
DP	10.1	11	21
	'Teddy'		
Proposed method	6.45	7.12	9.31
AdaptingBP	4.22	7.06	11.8
Double BP	5.53	8.30	9.63
Graph cut	12	17.6	22
DP	14	21.6	20.6
	'Cones'		
Proposed method	7.25	8.09	10.66
AdaptingBP	2.48	7.92	7.37
Double BP	2.90	8.78	7.79
Graph cut	4.89	11.8	12.1
DP	10.5	19.1	21.1

Table 1–Evaluation results based on the online Middlebury stereo benchmark system.

can be seen that the multiwavelet based algorithm produces the second best results for 'Cones' and 'Teddy' stereo test images, while for 'Tsukuba' and 'Venus' it ranks third and respectively third relative to the other four algorithms used for this comparison.

5. CONCLUSION

This paper introduced a hierarchical stereo correspondence matching technique based on multiwavelet transforms and global error energy minimization. The approximation subbands of the two views were used to generate a set of four disparity maps using a global error energy minimization algorithm. The resulting four disparity maps were then combined using a Fuzzy algorithm to form an initial low resolution disparity map, which was then refined by hierarchically propagating it to the finer levels. Results show that the proposed technique produces a disparity map with significantly less mismatch errors compared to applying the same GEEM algorithm in the spatial domain or in the wavelet transform domain. The performance of the proposed multiwavelet based algorithm has been compared to other well-known techniques benchmarked and published in the Middlebury database and the results indicate that the proposed multiwavelet based algorithm works well against many well established algorithms ranked at top positions in the Middlebury database. The multichannel nature of the multiwavelets and the different spectral content of the resulting subbands allow for greater correspondence matching flexibility than in the case of wavelets, and explain why the multiwavelet based technique performs better than when similar GEEM algorithms were applied in the wavelet and respectively the spatial domain, highlighting the potential of the multiwavelet transform in stereo correspondence applications.

REFERENCES

[1] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7-42, April 2002.

[2] Q. Jiang, J. J. Lee and M.H. Hayes, "A wavelet based stereo image coding algorithm," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 99),* vol.6, pp. 3157 - 3160, March 1999.

[3] G. Caspary and Y. Zeevi, "Wavelet-based multiresolution stereo vision," *Proc. IEEE International Conference on Pattern Recognition (ICPR 2002),* pp. 680- 683 , Dec. 2002.

[4] I. Sarkar and M. Bansal, "A wavelet-based multiresolution approach to solve the stereo correspondence problem using mutual information," *IEEE Transactions on System, Man, and Cybernetics*, vol. 37, pp. 1009-1014, August 2007.

[5] J. Li, H. Zhao, X. Zhou and C. Shi, "Robust stereo image matching using a two-dimensional monogenic wavelet transform," *Optics Letters*, vol. 34, pp. 3514-3516, 2009.

[6] S. Mallat, "A Wavelet Tour of Signal Processing," 3rd ed., Academic Press, 2009.

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

[8] 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.

[9] P. Bagheri Zadeh and C. Serdean, "Stereo Correspondence Matching: Balanced Multiwavelets versus unbalanced Multiwavelets," The 2010 European Signal Processing Conference (EUSIPCO-2010), pp. 1509-1513 , August 2010.

[10] V. Strela, "Multiwavelets: theory and applications," PhD thesis, MIT, 1996.

[11] 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.

[12] R.C. Gonzalez, R.E. Woods and S.L. Eddins, Digital image proc. second edition, Prentice Hall, pp. 75-142, 2002.

[13] http://vision.middlebury.edu/stereo/, last access 15 February 2011.

[14] A. Klaus, M. Sormann and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," *Proc. of Int. Conference on Pattern Recognition (ICPR 2006)*, vol. 3, pp. 15-18, Aug. 2006.

[15] Q. Yang, L. Wang, R. Yang, H. Stewenius and D. Nister, "Stereo Matching with Color-Weighted Correlation, Hierarchical Belief Propagation, and Occlusion Handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp.1-13, March 2009.

[16] V. Kolmogorov and R. Zabih , "Multi-camera scene reconstruction via graph cuts," *7th European Conference on Computer Vision (ECCV 2002)*, pp. 8-40, May 2002.

[17] D. Scharstein, "View Synthesis Using Stereo Vision," *Lecture Notes in Computer Science (LNCS),* vol. 1583, Springer-Verlag, 1999.