

## Image Matching using TI Multi-Wavelet Transform

Asim Bhatti, Saeid Nahavandi and Hong Zheng

CRC for Cast Metals Manufacturing (CAST), School of Engineering and  
Technology, Deakin University, Geelong 3217, Australia

[ambh@deakin.edu.au](mailto:ambh@deakin.edu.au) , [nahavand@deakin.edu.au](mailto:nahavand@deakin.edu.au) , [hzheng@deakin.edu.au](mailto:hzheng@deakin.edu.au)

**Abstract.** A multi-resolution image matching technique based on multi-wavelets followed by a *coarse to fine* strategy is presented. The technique addresses the estimation of optimal corresponding points and the corresponding disparity maps in the presence of occlusion, ambiguity and illuminative variations in the two perspective views taken by two different cameras or at different lighting conditions. The problem of occlusion and ambiguity is addressed by a geometric topological refining approach along with the uniqueness constraint whereas the illuminative variation is dealt by using windowed normalized correlation.

### 1 Introduction

The 3D reconstruction [4,8,9] process can be categorized into three main categories, calibration (calculating the intrinsic and extrinsic parameters of the camera) [1,3], finding the corresponding pairs of points projected from the same 3D point on to the two perspective views [2,5,6,7,10,18,19], and triangulation to project the 2D information back to the 3D space in order to create a 3D model [4,8,9,11]. The calibration and triangulation strategies are quite mature in both theoretical and applicative perspective but finding correct corresponding points from more than one perspective views still suffers from many problems like occlusion, ambiguity, illuminative variations and radial distortions, etc.

The motivation for using multi-wavelets (wavelets with more than one scaling and wavelet functions) is because of the fact that ever since their discovery, multi-wavelets have been the focus of a lot of research in signal processing and pure mathematics. The interest in multi-wavelets is mainly due to the fact that they produce promising results in many applications such as speech, image and video compression, denoising, communications, computer and machine vision [12,13,15-21]. Their success stems from the fact that they can simultaneously possess the good properties of orthogonality, symmetry, high approximation order and short support which is not possible in the

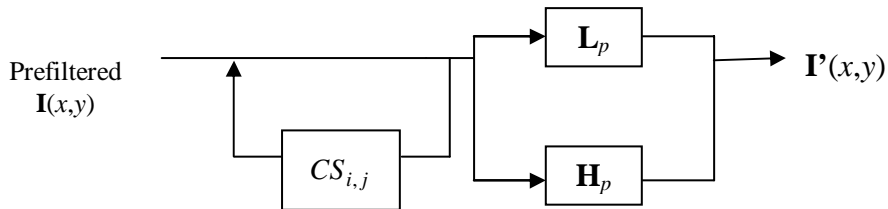
scalar case. Multi-wavelets already have been proven to perform better, than scalar wavelets, in applications like image compression and denoising [16,17,21] etc. In the application of correspondence matching some work has already been done using complex scalar wavelets [18,19] and convincing results have been achieved. As multi-wavelets have proven to perform better than scalar ones, due to their extra properties, there is a great deal of motivation to apply them in the application of correspondence matching. For that purpose a multi-resolution approach based on Multi-wavelets is used to decompose the images in order to perform the *coarse to fine* matching process.

The rest of the work is organized as follows: in the next section a brief explanation of achieving translation invariance from discrete dyadic wavelet transform presented by Mallat [12]. Section 3 is about the image matching algorithm proposed in that work with complete description of different parts involved in the algorithm. Some results of disparity maps along with the ground truth disparities are shown in section 4. Section 5 is about the conclusion followed by the references.

## 2 Need of Translation Invariance

The main drawback of the discrete dyadic wavelet or multi-wavelet transform is their shift and rotation variability [20,21], which are the common factors involved in the process of stereo vision. Translation and rotation variance means there is no direct relation between the wavelet coefficients of a transformed image and its translated or rotated versions. Thus using multi-resolution analysis without taking these factors in account will end up with catastrophic results and will be useless in the context of finding optimal corresponding points. The rotational variations can be dealt with physical alignment of the cameras along the base line (the line joining the two cameras) or more precisely by rectifying the two views. In order to deal with the translation variation a simplified procedure is performed to get Translation Invariant multi-wavelet (TIMW) coefficients from a non TI-MW transform introduced by many authors [20,21]. The TIMW transform used in that work is the modified version of V. Strela's technique [16]. The problem of translation variance arises due to the fact of factor-2 decimation involved in the decomposition process, and is solved by filtering 4 circular shifts of the preprocessed signal into the filter bank instead of one, for each level of decomposition. For the sake of minimizing the computational complexity of the matching process the details are

averaged over the 4 circular shifts after decomposition at each level. One level of TIMW decomposition is shown below



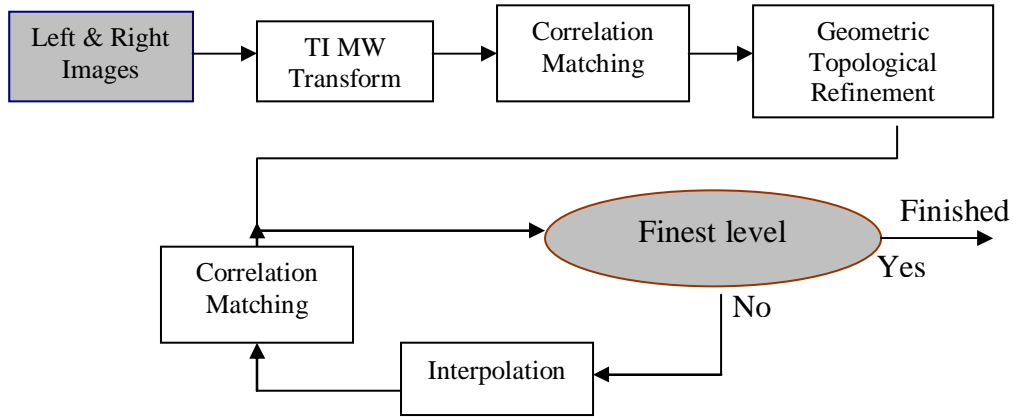
**Fig. 1** 1-level TI Multi-Wavelet transform

In Fig.1,  $CS_{i,j}$  represents the circular shifts with respect to  $i \in [0, 1]$  (row wise) and  $j \in [0, 1]$  (column wise) with 0 as no shift whereas 1 for one shift in respective direction.

### 3 Image Matching Algorithm

The first step of the matching process is the TIMW transform up to a desired level  $N$ , which is taken 4 in the proposed work, ends up with  $3Nr^2$  matrices of numbers which in fact are TIMW coefficients representing the details or discontinuities of the images at different resolutions respective to the decomposition level. Where  $r$  is the multiplicity of the multi-wavelets where as scalar wavelets has unit multiplicity. For example using “Chui-Lian” multi-wavelets [22] ends up with 12 whereas “mw112\_r3\_p2” [13] ends up with 27 TIMW coefficient matrices, due to the multiplicity 2 and 3 respectively. For simplicity, TIMW coefficient matrices will be denoted by  $CM_r$  and  $CM$  for reference and other image, respectively.

The TIMW decomposition is followed by a coarse to fine strategy, which involves the initial search of points at the coarsest level of decomposition and then their interpolation up to the finest level. In cases where more than one candidate matches, for any point in the reference space, a geometric topological refinement is performed to pick the optimal one. Uniqueness constraint is used, on the other hand, to choose the optimal one if more than one points in the reference space pairing with same point in the right. A block diagram representing the complete matching algorithm is shown below.



**Fig. 2.** Complete matching algorithm

### 3.1 Correlation

After TIMW transform, matching process starts at the coarsest level with the area based search, involving the calculation of normalized correlation (NC) for each coefficient of  $CM_r$  through CM. Instead of single coefficients a window of  $(2n \times 1, 2n \times 1)$ ,  $n$  is usually taken within the range of [3 5], is used centered at the coefficient under consideration. The NC score is defined as

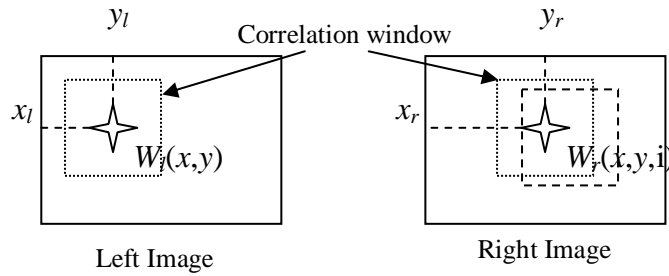
$$CS_{N,k}(x, y) = \frac{\sum_{i,j} [(W_{r,N,k} - \overline{W_{r,N,k}}) \times (W_{N,k} - \overline{W_{N,k}})]}{\sqrt{\sum_{i,j} [W_{r,N,k} - \overline{W_{r,N,k}}]^2} \times \sqrt{\sum_{i,j} [W_{N,k} - \overline{W_{N,k}}]^2}} \quad (1)$$

where  $W_{r,N,k} = W_{r,N,k}(x+i, y+j)$ ,  $W_{N,k} = W_{N,k}(x+i+dx, y+j+dy)$ , and  $W_{r,N,k}(x, y)$  represents  $(2n \times 1, 2n \times 1)$  window of coefficients of  $k^{th}$   $CM_r$  and  $N^{th}$  decomposition level where as  $d_x, d_y$  represents the disparity values in  $x$  and  $y$  directions respectively, where as

$$\overline{W_{N,k}(x, y)} = \sum_x \sum_y W_l(x, y) / (2n+1)^2 \quad (2)$$

The main purpose of subtracting the averages is to minimize the effect of illuminative variations between the two images. Another good

feature of that correlation expression is that it is invariant to the changes from  $W_1$ ,  $W_2$  to  $a_1W_1 + b_1$ ,  $a_2W_2 + b_2$  respectively [10]. The correlation process can be better visualized by Fig 3.



**Fig. 3.** Windowed normalized correlation process

As it is quite obvious from (1) that the values of  $CS_{N,k}(x, y, d)$  lie within  $[-1 \ 1]$ , with  $-1$  for anti-correlated and  $1$  for the identical correlation windows. The coefficients with the maximum correlation score, from each CM, will be taken as the candidate match, having a set of at most  $3r^2$  candidates for each coefficient in the  $CM_r$ . A constraint is then applied to select the most consistent matches. A coefficient with correlation score higher than a predefined threshold  $t$  will be selected as a candidate match and will be used for further processing. The threshold is usually taken within the range  $[0.5 \ 0.8]$ .

### 3.2 Geometric Topological Refining

After the correlation process, each coefficient of  $CM_r$  will have a set of candidate matches. These candidate matches can either be pointing to the same location or different in CM and vice versa. Here an assumption is used, which gives some reference locations for further refinement.

*If all the candidates are pointing to the same location and the size of the set of candidates is bigger than or equal to  $3r^2/2$ , that coefficient will be considered as true match.*

In order to find an optimal one from the set of candidates, containing different locations, geometric features like *relative distances* and *angles* (slopes of lines for simplicity) are calculated, which are the invariant features through many geometric transformations like Euclidean, metric, etc. The occurrence of these transformations is very common in the applications of stereo vision and 3D modeling, which is the ultimate goal of that work. The candidate having closest geometric topology w.r.t

the coefficient in  $CM_r$  will be counted an optimal match, while considering true matches as a reference. In other words candidate with highest match strength among all others will be selected as an optimal match. The match strength is defined as

$$MS_k(x, y) = \overline{CS_k} \left( \frac{e^{-rdd_k} + e^{-spd_k}}{2} \right) \quad (3)$$

where  $k$  is the number of candidate matches for a specific coefficient in  $CM_r$  and  $\overline{CS_k}$  is the average correlation score of  $k^{th}$  candidate defined as

$$\overline{CS_k} = \sum_k CS_k(x, y, d) / 3r^2 \quad (4)$$

$rdd_k$  is the average relative distance difference between the coefficient in  $CM_r$  and its  $k^{th}$  candidate with reference to the true matches, related to both left and right images, as given in (5).

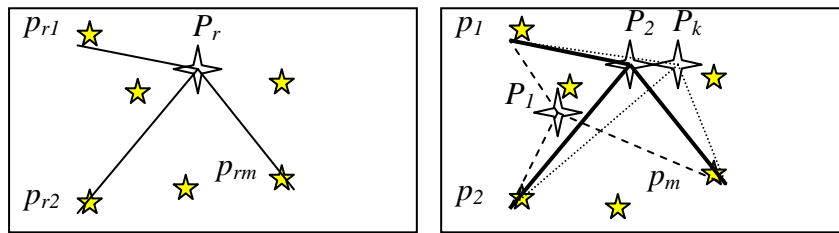
$$rdd_k = \sum_{i=1}^m \frac{|d(P_r, p_{ri}) - d(P_k, p_i)|}{d(P_r, p_{ri}) + d(P_k, p_i)} / m \quad (5)$$

where  $d(P_r, p_{ri})$  and  $d(P, p_i)$  are the Euclidian distances between points  $P_r$  and  $p_{ri}$  in the reference space and points  $P$  and  $p_i$  in the other space respectively, where as  $p_{ri}$  and  $p_i$  are the true matches. Similarly  $spd_k$  is the relative slope difference between the coefficient in  $CM_r$  and its  $k^{th}$  candidate with reference to the true matches and is defined as

$$spd_k = \sum_{i=1}^m \frac{|sp(P_r, p_{ri}) - sp(P_k, p_i)|}{sp(P_r, p_{ri}) + sp(P_k, p_i)} / m \quad (6)$$

where  $sp(P, p_i) = \frac{y_P - y_{p_i}}{x_P - x_{p_i}}$ .

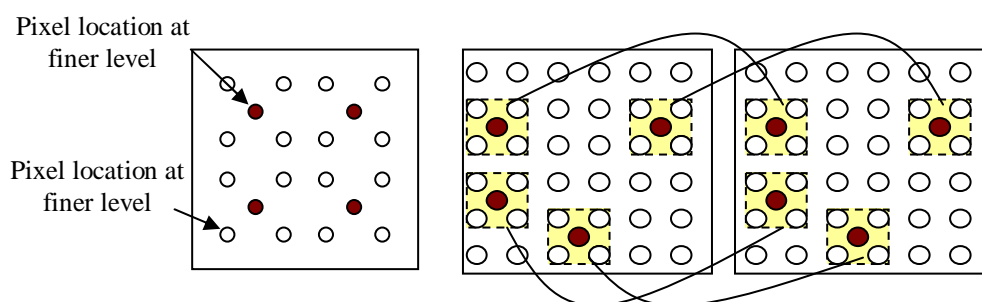
In order to minimize the effect of a wrong true match, chosen after the correlation step,  $rdd_k$  and  $spd_k$  are calculated  $m$  times, instead of one, by picking a random point every time from the bin of true matches and averaging over  $m$ . The complete geometric topological approach can be visualized by a diagram shown below



**Fig. 4.** Geometric topological refining technique

### 3.3 Interpolation

The matching process at the coarsest level ends up with a number of matching pairs which needs to be interpolated to the finer level. The constellation relation between the coefficients at coarser and finer levels can be visualized by taking the decimation of factor 2 into consideration as can be seen in Fig 5. From this constellation, it is quite clear that each pixel at coarser level represents 4 pixels at finer scale. At that stage (finer) only those sets of pixels will be considered, which has their corresponding matched pairs at the coarser level as shown in figure below with shaded area. After the matches are interpolated to the finer level correlation process is performed again to check the consistency of the matches. The matches fulfilling the criterion will be taken as credible matches for further processing whereas rest of the matches are discarded, hence refining the matches up to the finest level and leaving most consistent matches at the end of the process.



**Fig. 5.** Constellation Relation between coarser and finer level coefficients (left), Interpolated coefficient locations at finer level (right)

## 4 Results

As there is no independent and absolute way of checking the performance of the matching algorithm, a comparative performance check based on the calculation of first order re-projection error, which involves the calculation of fundamental matrix [2, 14] is performed and is given in Table 1.

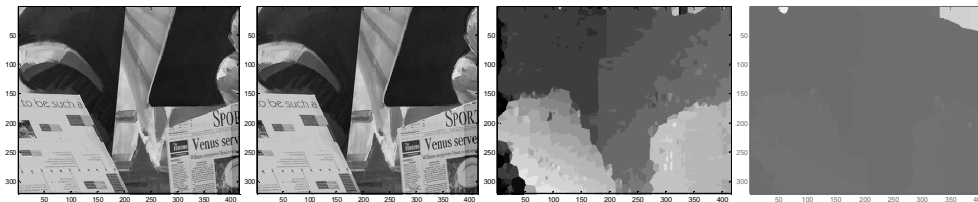
**Table 1.** Statistical Comparison of matching processes

Tool	Images	matches	Error [min max]	Variance of error
M-wavelets	Venus(320×416)	22915	[1.0487e-35 7.7637e-28]	4.0181e-57
	Pentagon(512×512)	48525	[2.1098e-33 2.3957e-28]	9.1162e-57
wavelets	Venus(320×416)	19595	[1.6673e-33 1.1419e-27]	2.1950e-56
	Pentagon(512×512)	37242	[2.5148e-35 8.8560e-28]	3.0770e-56
Torr-tool	Venus(320×416)	1000	[1.8342e-29 0.0039]	6.2411e-08
	Pentagon(512×512)	999	[8.3120e-029 0.5084]	0.0010

Three different statistics are shown in table.1, obtained by applying proposed algorithm on two different pair of images shown in Fig.6, using Daubechies's wavelet [15] and mw112\_r3\_p2 MW [13], whereas the last two rows of the table are the results of the technique presented by P. Torr in [14] which involves feature extraction followed by correlation matching and has been fixed to find 1000 best features. Accordingly the calculated disparity maps of two different image pairs are also shown. First pair of images is taken from <http://vasc.ri.cmu.edu/idb/images/ster-eo/pentagon> which is an aerial image pair of famous "Pentagon" building and second image pair is known by the name Venus and is taken from <http://cat.middlebury.edu/stereo/data.html>. Reasonably promising results are obtained in both cases. Due to the space shortage ground truth disparities are not shown and can be found from their respective websites. No special procedure, except linear interpolation, is performed to fill the gap between the matched points, which can be improved by applying the techniques like [23].







**Fig. 6.** *Top 1-2:* Stereo pair of Pentagon images, *Top 3:* MW disparity map, *Top 4:* Wavelet disparity map, *Bottom 1-2:* Stereo pair of Venus images *Bottom 3:* MW disparity map, *Bottom 4:* Wavelet disparity map. (1-4 is from left to right).

Multi-wavelets have performed well in criteria like re-projection error and producing disparity maps while applying to completely two different categories of images. On the other hand the performance of Daubechies' Wavelets is reasonably well in terms of calculating re-projection error but lost the track especially while calculating the disparity map of Venus image as can be seen in Fig 6 (*bottom 4*).

## 5 Conclusion

A multi-resolution image matching technique based on TIMW transform is presented. Multi-wavelets have performed well and proved to have the potential, as a good tool, for solving the problems of finding optimal corresponding points. It's also shown that multi-wavelets perform better than scalar ones as it has been shown by many authors in many other applications like compression, denoising, etc.

A geometric topological refining approach is also presented which is quite useful and has performed well in finding the optimal corresponding points even in the presence of occlusion, ambiguity and illuminative variations, which are few of the major problems involved in the stereo vision applications.

## Reference

1. Z. Zhang. "A flexible new technique for camera calibration". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330-1334, 2000.
2. Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", Research Report 2273, INIRA Sophia-Antipolis, France, May 1994. submitted to *Artificial Intelligence Journal*
3. J. Heikkilä, & O. Silvén, "A Four-step Camera Calibration Procedure with Implicit Image Correction". *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, p. 1106-1112. 1997.
4. T. J. Moyung. Incremental 3D reconstruction using stereo image sequences. PhD thesis, Waterloo, Canada, 2000

5. Y. Boykov, O. Veksler, and R. Zabih, "A variable window approach to early vision," *IEEE transaction on Pattern Analysis and Machine Intelligence*, vol. 20, December 1998.
6. W. H. Liao and J. K. Aggarwal, "Cooperative matching paradigm for the analysis of stereo image sequences," *Int. J. of Imaging Systems and Technology*, 9(3):192-200, 1998.
7. D. Chetverikov and J. Verest'oy, "Tracking feature points: a new algorithm," *Proc. Int. Conf. on Pattern Recognition*, 1436-1438, 1998.
8. E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.
9. A. Faugeras, "Three dimensional Computer Vision - A geometric Approach Viewpoint. MIT Press, 1993.
10. Olivier Faugeras, Bernard Hotz, Hervé Mathieu, Thierry Viéville, Zhengyou Zhang, Pascal Fua, Eric Thérion, Laurent Moll, Gérard Berry, Jean Vuillemin, Patrice Bertin, Catherine Proy, Real time correlation-based stereo: algorithm, implementations and application, *INRIA Research Report 2013*, August 1993..
11. F. Kahl, R. Hartley, and K. Astrom, "Critical configurations for n-view projective reconstruction", In *Proc. Conf. Computer Vision and Pattern Recognition, Hawaii, USA*, 2001.
12. S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans. PAMI* 11 (1989) 674-693.
13. A. Bhatti, H. Özkaramanli, "M-Band multiwavelets from spline super functions with approximation order", *International Conference on Acoustics Speech and Signal Processing, ICASSP, Orlando, FL, USA*, May 2002.
14. P. H. S. Torr, "A Structure and Motion Toolkit in Matlab, Interactive Adventures in S and M, Microsoft, Technical Report MSR-TR-2002-56, June 2002.
15. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
16. V. Strela, P.N. Heller, G. Strang, P. Topiwala and C. Heil, "The application of multi-wavelet filter banks to signal and image processing," *IEEE Trans. on Image Processing*, 8(4):548-563, April 1999.
17. V. Strela and A.T. Walden, "Orthogonal and biorthogonal multiwavelets for signal denoising and image compression," in *Wavelet applications, Proceedings-SPIE, The International Society for Optical Engineering*, 5<sup>th</sup> Meeting, Orlando; FL, vol. 3391, pp 96-107, April 1998.
18. He-Ping Pan, "Uniform full-information image matching using complex conjugate wavelet pyramids", *XVIII ISPRS Congress, Vienna, 1996. International Archives of Photogrammetry and Remote Sensing*, Vol.XXXI.
19. J. F. A. Margary and A. dick., "Multiresolution stereo image matching using complex wavelets", In *Proc. 14<sup>th</sup> Int. Conf. on Pattern Recognition (ICPR)*, volume I, pages 4-7, August 1998.
20. I. Cohen, S. Raz and D. Malah, "Adaptive time-frequency distributions via the shiftinvariant wavelet packet decomposition", *Proc. of the 4th IEEE-SP Int. Symposium on Time-Frequency and Time-Scale Analysis*, Pittsburgh, Pennsylvania, Oct. 1998.
21. R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising", in: A. Antoniadis and G. Oppenheim, ed., *Wavelet and Statistics*, Lecture Notes in Statistics, Springer-Verlag, 1995, pp. 125-150.
22. C. K. Chui and J. Lian, "A Study of Orthonormal Multi-wavelets," *J. Applied Numerical Math.* Vol.20, no.3, pp273-298, 1996.
23. D. Scharstein and R. Szeliski, "a taxonomy and evaluation of dense two-frame stereo correspondence algorithms, *IJCV*, Vol 47,no 1, pp 7-42, May 2002.