This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30005260

Every reasonable effort has been made to ensure that permission has been obtained for items included in Deakin Research Online. If you believe that your rights have been infringed by this repository, please contact drosupport@deakin.edu.au

Copyright : 2003, The Authors
Disparity Estimation 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

E-mail: ambh, nahavand, hzheng@deakin.edu.au

Abstract.

A multi-resolution image matching technique based on translation invariant discrete multi-wavelet transform followed by a coarse to fine matching 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 explicitly by a geometric optimization approach along with the uniqueness constraint whereas the illuminative variation is dealt with by using windowed normalized correlation on the discrete multi-wavelet coefficients.

Keywords: TI DMWT, normalized correlation, geometric optimization, disparity.

1. Introduction

The 3D reconstruction process can be categorized into three main categories, calibration (calculating the intrinsic and extrinsic parameters of the camera) [2,3,5,6], finding the corresponding pairs of points projected from the same 3D point on to more than one perspective views [13,4,8,9,10,24,25,26], and triangulation to project the 2D information back to the 3D space in order to create a 3D model [1,7,11,12,14]. 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 due to many problems like occlusion, ambiguity, illuminative variations, radial distortions, etc.

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 motivation for using multi-wavelets is because of the fact that ever since their discovery, multi-wavelets (wavelets with more than one scaling and wavelet functions) have been the focus of a lot of research in signal processing and is mainly due to the fact that they produce promising results in many applications such as speech, image and video compression, denoising, computer and machine vision [15,16,17,21,22,24,25,28,30]. Their success stems from the fact that they can simultaneously posses 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 [15,18,21,22,23,28,30,31,32] etc. In the application of correspondence matching some work has already been done using complex scalar wavelets [24-27]. In [24,25] He-Ping used complex scalar wavelets to find the corresponding points in the aerial images in order to create 3D model and convincing results have achieved. He also discussed in [25] many properties that wavelets should posses to show even better performance in the context of correspondence matching. Multi-wavelets, on the other hand, already posses almost all of the properties mentioned in [25], simultaneously, except a few like translation invariance which will be taken care by a simple method given in section 2. 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.
The rest of the work is organized as follows: in the next session a brief introduction of multi-wavelet theory. A brief explanation of the lack of translation invariance involved in the discrete dyadic filter tree [15] and its effect on applications like correspondence matching. 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. Multi-wavelets and translation invariance

Multi-resolution can be generated not just in the scalar context i.e., with just one scaling function and one wavelet but also in the vector case where there is more than one scaling functions and wavelets. The latter case leads to the notion of multi-wavelets. A multi-wavelet basis is characterized by more than one scaling function and wavelets. Here \( r \) denotes the multiplicity in the vector setting with \( r > 1 \). Multi-scaling functions satisfy the matrix dilation equation or the refinement equation

\[
\Phi(t) = \sum_{k} C_k \Phi(2r - k) \quad (1)
\]

Similarly for the multi-wavelets the matrix dilation equation is given by

\[
\Psi(t) = \sum_{k} W_k \Phi(2r - k) \quad (2)
\]

where \( \Phi(t) = [\phi_0(t), \phi_1(t), \ldots, \phi_{r-1}(t)]^T \), \( \Psi(t) = [\psi_0(t), \psi_1(t), \ldots, \psi_{r-1}(t)]^T \) and \( C_k \) and \( W_k \) are real \( r \) by \( r \) matrices of multi-filter coefficients that iteratively and with rescaling define the scaling and wavelet functions respectively i.e., the link between functions and filters is iteration with rescaling. For more information, about the generation and applications of multi-wavelets with, desired approximation order and orthogonality, the interested readers are referred to [15, 16, 17, 20, 22, 33].

As the filter coefficients are matrices instead of scalars, multiple streams of inputs must be fed to the filter bank depending on the multiplicity of the filters. To cope with the obligation of multiple input streams, identity prefilter is used which is also known as repeated row preprocess in the literature [21, 22]. An example of discrete Multi-wavelet transform (DMWT) using famous GHM multi-wavelet [33] is shown in Figure 1 which is applied on the Lena image. In Figure 1 (a) \( LH, HL, HH \) represents the horizontal, vertical and diagonal details of the Lena image, respectively.

The DMWT (Mallat’s dyadic filter tree [15]) has established, on one hand, an impressive reputation as a tool in many image processing applications like compression, denoising etc. but on the other hand suffer due to the lack of translation and rotation invariance [28, 29, 30], which is the common factor involved in many applications like stereo vision, etc. Lack of translation invariance means a small shift in the samples of input will cause a great change in the DMWT coefficients. Thus using DMWT 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) and more precisely by rectifying the two views. The problem of translation variance arises due to the fact of factor-2 decimation involved in the decomposition process. A well known way of providing translation invariance is to use the dyadic filter tree without the decimation operator, but this suffers from high redundancy in the output information and high computational cost in the following processes. In order to deal with the translation variation a simplified procedure is performed to get Translation Invariant DMWT (TIDMWT) coefficients from a non TI-DMWT introduced by many authors [28,29,30]. For that purpose a modified and updated version of V. Strela’s technique [21] is used which involves the filtering of 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. A simplified block diagram representing the above mentioned procedure is shown in Figure 2, where \( CS(i,j) \) represents the circular shifts with respect to \( i \in [0 \, 1] \) (row wise) and \( j \in [0 \, 1] \) (column wise) with 0 as none whereas 1 for one shift in respective direction.

3. Correspondence matching

In the proposed work, matching process is performed on the pair of rectified images. After taking TIDMWT up to level \( N \), there will be \( 3N^2 \) TIDMWT coefficient matrices, which will take part in the matching process. Where \( r \) represents the multiplicity of the multi-filters \( C_k \) and \( W_k \). In order
to find the initial matches normalized correlation is performed at the coarsest level TIDMWT coefficients as shown below

$$\text{NCS}_{N,k,d} = \frac{\sum_{n=1}^{2^{n+1}} [NW_{1,N,k} \times NW_{2,N,k,d}]^2}{\sqrt{\sum_{n=1}^{2^{n+1}} NW_{1,N,k}^2} \times \sqrt{\sum_{n=1}^{2^{n+1}} W_{2,N,k,d}^2}}$$

where $NW_{1,N,k}$ represents normalized coefficient matrix of size $2n \times 1$ taken from $k^{th}$ TIDMWT coefficient matrix at $N^{th}$ level of decomposition, related to first (left) and second images and can be expressed explicitly as

$$NW_{1,N,k} = W_{1,N,k} - \bar{W}_{1,N,k}$$

where $\bar{W}_{1,N,k}$ is the mean value of coefficient matrix where as $d$ in (3) represents the disparity value. It’s quite obvious from the expression of (3) that NCS (Normalized Correlation Score) lies within [-1 1]. A threshold is then applied to select the most consistent matches.

### 3.1 Geometric optimization

The completion of correlation process ends up with a number of corresponding pairs of points. Some points will have more than one candidate pairs satisfying the above criteria and need to be further examined to find the optimal one. Uniqueness restrain is applied if there are more than one candidate in the reference (left) TIDMWT coefficient matrix pairing with a single coefficient in the other matrix, by keeping the one with highest NCS value. To find the optimal one, if the above problem is in opposite order i.e. more than one candidate pairs in right TIDMWT coefficient matrix for some coefficient in the left, a geometric optimization is performed. As there are $3N^2$ coefficient matrices as a search space, the corresponding pairs which were consistent in pairing with same coefficients through more than half of the search space, i.e. paired to a single and same coefficient from more than $3N^2/2$ matrices, will be denoted as most consistent pairs (MCPs) and will be taken as references to find the optimal ones for the points having more than one candidates. For that purpose two geometric features, relative distances and slopes of lines joining the two points, are calculated for all, reference and candidate pairs, with respect to the MCPs. These features are invariant through many geometric transformations like affine, metric, etc. The occurrence of these transformations is very common in the applications of stereo vision and 3D modeling. The candidate having closest geometric topology to the topology of reference point will be counted as optimal match. In other words candidate with highest pair strength (PS) among all others will be selected as an optimal pair and is defined as

$$PS_m = \frac{\overline{NCS}_m(e^{-\overline{rdd}_{k}} + e^{-\overline{spd}_{k}})}{2}$$

where $\overline{NCS}_m$ is the averaged normalized correlation score for $m^{th}$ candidate if it paired more than once to some point and is defined as

$$\overline{NCS}_m = \sum_{mk} NCS_{N,k,m} / 3r^2$$

where as $\overline{rdd}_{k}$ and $\overline{spd}_{k}$ are the average of relative distance (Euclidean) differences and slopes, respectively. Obviously the candidate with higher NCS and lower geometric topological difference will be selected. To better visualize the concept, a simple diagram is shown in Figure 4.

### 3.2 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 Figure 4. After the matches are interpolated to the finer level correlation process is performed again, only for the locations having their corresponding pairs at the coarser level. The matches fulfilling the criterion will be taken as credible matches for next interpolation 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. A block diagram representing the complete process is shown in Figure 3.

### 4. Results

The algorithm proposed in that work is applied to a number of different kinds of real images and reasonably satisfactory results are obtained. For that purpose Mw112_r3_p2 multi-wavelet [17] is used which possess approximation order 2 and multiplicity 3. Three of the obtained results from completely different kind of images are shown here.

The first pair of ball images is taken from CSIRO’s website http://extra.cmis.csiro.au/IA/changs/stereo and as the ground truth was not available the disparity map
obtained from their algorithm is shown with the calculated disparity map, in Figure 5.

The second set of experiment is performed on the stereo pair of “Saw tooth” image, which is quite famous in the literature of stereo vision, and is taken from http://cat.middlebury.edu/stereo/data.html. As for comparison the ground truth disparity map is shown along with ground truth disparity in Figure 6. Figure 7 is about the calculated disparity map of famous “Pentagon” stereo pair of images and is taken from http://vasc.ri.cmu.edu/idb/images/stereo/pentagon.

5. Conclusion and future work

A multi-resolution image matching technique based on TIDMW transform is presented and quite promising results have achieved, a few of which are shown here. Multi-wavelets have performed well and proved to have the potential, as a good tool, for solving the problems of finding optimal corresponding points. A geometric optimization 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.

A few things that need to be taken care of as further improvement in the proposed work are the minimization of the computational cost in order to bring it close to the real time process and second is the handling of boundaries while creating disparity maps, which can be done by introducing the information of edges.

Reference


Figure 1: 1-level DMWT, left: original image, center: preprocessed multi-stream input, right: 1 level decomposed image

Figure 2: 1-level TI Multi-Wavelet transform

Figure 3: Block diagram of correspondence matching algorithm
Figure 4: Geometric optimization

Figure 5: Top: Stereo pair of original images, Bottom left: (calculated) Disparity map obtained from CSIRO website
Bottom right: Calculated disparity map
Figure 6: **Top:** Stereo pair of original Saw Tooth images, **Bottom left:** Ground Truth Disparity Map, **Bottom right:** Calculated disparity map

Figure 7: **Top:** Stereo pair of original Saw Tooth images, **Bottom:** Calculated disparity map