2

Stereo Correspondence Estimation based on Wavelets and Multiwavelets Analysis

Asim Bhatti and Saeid Nahavandi
Intelligent Systems Research Lab., Deakin University
Australia

1. Introduction

Stereo vision has long been studied and a number of techniques have been proposed to estimate the disparities and 3D depth of the objects in the scene. This is done by estimating the pixel correspondences of similar features in the stereo perspective views originated from the same 3D scene (O. Faugeras 2001; R. Hartley 2003). However, finding correct corresponding points is subject to a number of potential problems like occlusion, ambiguity, illuminative variations and radial distortions (D. Scharstein 2002). A number of algorithms have been proposed to address some of the aforementioned issues in stereo vision however it is still relatively an open problem.

Current research in stereo vision has attracted a lot of focus on multiresolution techniques, based on wavelets/multiwavelets scale-space representation and analysis, for correspondence estimation (S. Mallat 1993; He-Ping Pan 1996). However, very little work has been reported in this regard. The main advantage of these algorithms is their hierarchical nature that exhibit behaviour similar to iterative optimization algorithms. These algorithms generally operate on an image pyramid where results from coarser levels are used to constrain regional search at finer levels. The coarse-to-fine techniques adopted in these algorithms are considered to be a middle approach unlike other existing algorithms where correspondences are established based either purely on local search (L. Di Stefano 2004; Q`ıngxiong Yang 2006) or as a global cost-function optimization problem (D. Scharstein 1998; M. Lin 2003) with their respective shortcomings.

In this work, earlier contributions regarding the application of wavelet/multiwavelets in stereo vision is presented highlighting shortcomings, involved in those techniques such as translation and rotational variance (R. R. Coifman 1995; S. Mallat 1999) inherited in the discrete wavelet/multiwavelet transformation. Furthermore, correspondence estimation, directly using wavelet coefficients for aggregation costs, is very sensitive to noise and illuminative variation that generally exists in between the stereo perspective views.

In addition, a new novel and robust algorithm is presented (Asim Bhatti 2008) using the hierarchical correspondence estimation approach where wavelets/multiwavelets transform modulus maxima (*WTMM*) are considered as corresponding features, addressing the issue of translation invariance. *WTMM* defines translation invariant features with phases pointing normal to the surface. The proposed algorithm introduced a new comprehensive selection criterion called *strength of the candidate* (*SC*) unlike many existing algorithms where selection

Source: Stereo Vision, Book edited by: Dr. Asim Bhatti,
ISBN 978-953-7619-22-0, pp. 372, November 2008, I-Tech, Vienna, Austria

is solely based on different aggregation costs. The *SC* involves the contribution of probabilistic weighted normalized correlation, symbolic tagging and geometric topological refinement. Probabilistic weighting involves the contribution of more than one search spaces especially in the case of multi-wavelet based multi-resolution analysis. Symbolic tagging procedure helps to keep the track of different candidates to be an optimal candidate. Furthermore, geometric topological refinement addresses the problem of ambiguity due to geometric transformations and distortions that could exist between the perspective views. The geometric features used in the geometric refinement procedure are carefully chosen to be invariant through many geometric transformations such as affine, metric and projective (M. Pollefeys 2000).

The developed vision system based on the proposed algorithm is very simple and cost effective as it consists of only a stereo cameras pair and a simple fluorescent light. The developed system is capable of estimating surface depths within the tolerance of 0.5 mm. Moreover, the system is invariant to illuminative variations and orientation of the objects in the input image space, which makes the developed system highly robust. Due to its hardware simplicity and robustness, it can be implemented in different factory environments without a significant change in the setup.

2. Wavelets analysis in stereo vision: a background

Wavelet theory has been explored very little up to now in the context of stereo vision. Some work has been reported in applying wavelet theory for addressing correspondence problem. To the best of author's knowledge, Mallat (S. Mallat 1991; S. Mallat 1993) was the first who used wavelet theory concept for image matching by using the *zero-crossings* of the wavelet transform to seek correspondence in image pairs. In (S. Mallat 1993) Mallat also explored the signal decomposition into linear waveforms and signal energy distribution in time-frequency plane. Afterwards, Unser (M. Unser 1993) used the concept of multi-resolution (*coarse to fine*) for image pattern registration using orthogonal wavelet pyramids with *spline bases*. Olive-Deubler-Boulin (J. C. Olive 1994) introduced a block matching method using orthogonal wavelet transform whereas (X. Zhou 1994) performed image matching using orthogonal Haar wavelet bases. Haar wavelet bases are one of the first and simplest wavelet basis and posses very basic properties in terms of smoothness, approximation order (A. Haar 1910), therefore are not well adapted for most of the imaging applications, especially correspondence estimation problem.

A more comprehensive use of wavelet theory based multi-resolution analysis for image matching was done by He-Pan in 1996 (He-Ping Pan 1996; He-Ping Pan 1996). He took the application of wavelet theory a bit further by introducing a complete stereo image matching algorithm using complex wavelet basis. In (He-Ping Pan 1996) he explored many different properties of wavelet basis that can be well suited and adaptive to the stereo matching problem. A number of real and complex wavelet bases were used and transform is performed using wavelet pyramid, commonly known by the name *Mallat's dyadic wavelet filter tree* (*MDWFT*) (S. Mallat 1999). The common problem with *MDWFT* is the lack of translation and rotation variance (R. R. Coifman 1995; I. Cohen 1998) especially in real valued wavelet basis. Furthermore similarity measures were applied on individual wavelet coefficients which are very sensitive to noise. Similarly, Magarey (J. Magarey 1998; J. Margary 1998) introduced algorithms for motion estimation and image matching, respectively, using complex discrete *Gabor-like* quadrature mirror filters. Afterwards, Shi

(Fangmin Shi 2001) applied *sum of squared difference* (*SSD*) technique on wavelet coefficients. He uses translation invariant wavelet transformation for matching purposes, which is a step forward in the context of stereo vision and applications of wavelet.

More to the wavelet theory, multi-wavelet theory evolved (G. Plonka 1998) in early 1990s from wavelet theory and enhanced for more than a decade. Their success, over scalar wavelet bases, 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 (Özkaramanli H. 2001; A. Bhatti 2002). Being a new theoretical evolution, multi-wavelets are still new and are not yet applied in many applications. In this work we will devise a new and generalized correspondence estimation technique based wavelets and multiwavelets analysis to provide a framework for further research in this particular context.

3. Wavelet and multiwavelets fundamentals

Before proceeding to the correspondence estimation algorithms it seems wise to provide brief background of wavelets and multiwavelets theory as the algorithm presented is based heavily on this theory. Furthermore this background will assist the user to understand the features that are used to establish correspondences between the stereo pair of images.

Classical wavelet theory is based on the dilation equations as

$$\phi(t) = \sum_h c_h \, \phi(Mt - h) \tag{1}$$

$$\psi(t) = \sum_h w_h \, \phi(Mt - h) \tag{2}$$

where c_h and w_h represents the scaling and wavelet coefficients whereas M represents the band of filter bank (A. Bhatti 2002). In addition, multiresolution 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 function and wavelet are involved. The latter case leads to the notion of multiwavelets. Multiwavelets bases are characterized by r scaling functions and r wavelets in contrast with one scaling function and one wavelet, i.e. r =1. Here r denotes the multiplicity in the vector setting with $r > 1$.

In the case of multiwavelets, scaling functions satisfy the matrix dilation equation as

$$\Phi(t) = \sum_h C_h \Phi(Mt - h) \tag{3}$$

Similarly for the multiwavelets the matrix dilation equation can be expressed as

$$\Psi(t) = \sum_h W_h \Phi(Mt - h) \tag{4}$$

In equations (3) and (4), c_h and w_h are real matrices of multi-filter coefficients whereas $\Phi(t)$ and $\Psi(t)$ can be expresses in terms of r scaling functions and r wavelets as

$$\mathbf{\Phi}(t) = \begin{bmatrix} \phi_0(t) \\ \phi_1(t) \\ \vdots \\ \phi_{r-1}(t) \end{bmatrix} \tag{5}$$

And

$$\psi(t) = \begin{bmatrix} \psi_0(t) \\ \psi_1(t) \\ \vdots \\ \psi_{r-1}(t) \end{bmatrix} \tag{6}$$

Generally only two band multiwavelets, i.e. $M = 2$, defining equal number of wavelets as scaling functions are used for simplicity. For further information about the generation and applications of multiwavelets, with desired approximation order and orthogonality, interested readers are referred to (S. Mallat 1999; A. Bhatti 2002).

3.1 Wavelet filter banks

Wavelet transformation produces scale-space representation of the input signal by generating scaled version of the approximation space and the detail space possessing the property

$$A_{s-1} = A_s \bigoplus D_s \tag{7}$$

where A_s and D_s represents approximation and detail space at lower resolution/scale and by direct sum constitutes the higher scale space A_{s-1}. In other words A_s and D_s are the subspaces of A_{s-1}. Expression (7) can be better visualized by Figure 1.

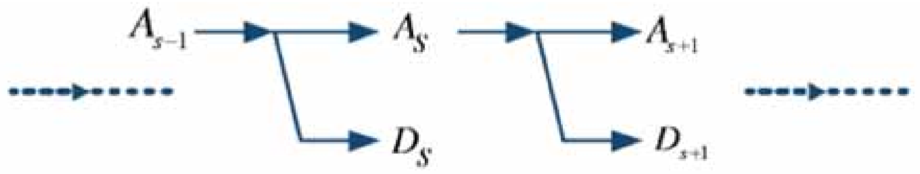

Fig. 1. wavelet theory based Multiresolution analysis

The use of Mallat's dyadic filter-bank (S. Mallat 1999) results in three different detail space components that are the horizontal, vertical and diagonal. Figure 2 can best visualize the graphical representation of the used filter-bank, where and W represents the low-pass and high-pass filters consisting of the scaling functions and wavelets coefficients, respectively.

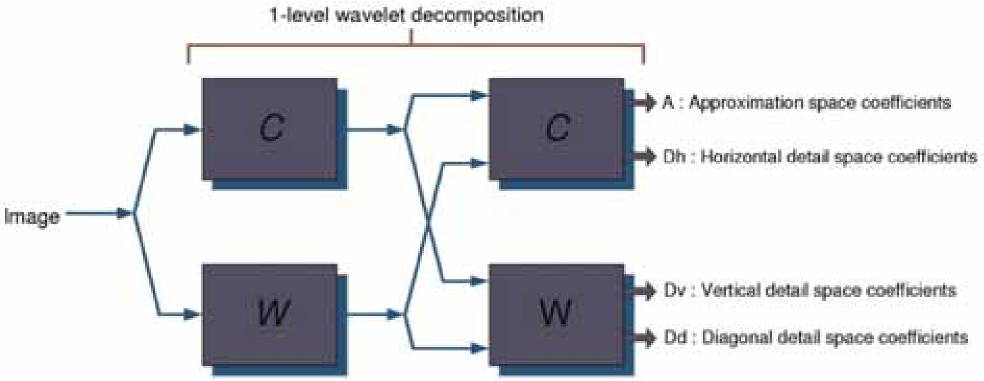
1-level wavelet decomposition
Image
C
W
C
W
A : Approximation space coefficients
Dh : Horizontal detail space coefficients
Dv : Vertical detail space coefficients
Dd : Diagonal detail space coefficients

Fig. 2. Mallat's dyadic wavelet filter bank

3.2 Wavelet transform modulus

The *Wavelet Transform Modulus* (*WTM*), in general vector representation, can be expressed as

$$WTM_{s,k} = W_{s,k} \angle \Theta_{W_{s,k}} \tag{8}$$

Where $W_{s,k}$ is

$$W_{s,k} = \sqrt{\left|D_{h,s,k}\right|^2 + \left|D_{v,s,k}\right|^2} \tag{9}$$

where $D_{h,s,k}$ and $D_{v,s,k}$ are the kth horizontal and vertical detail components at scale s. Furthermore $\Theta_{W_{s,k}}$ can be expressed as

$$\Theta_{W,s,k} = \begin{cases} \alpha(s,k) & if \quad D_{h,s,k} > 0 \\ \pi - \alpha(s,k) & if \quad D_{h,s,k} < 0 \end{cases} \tag{10}$$

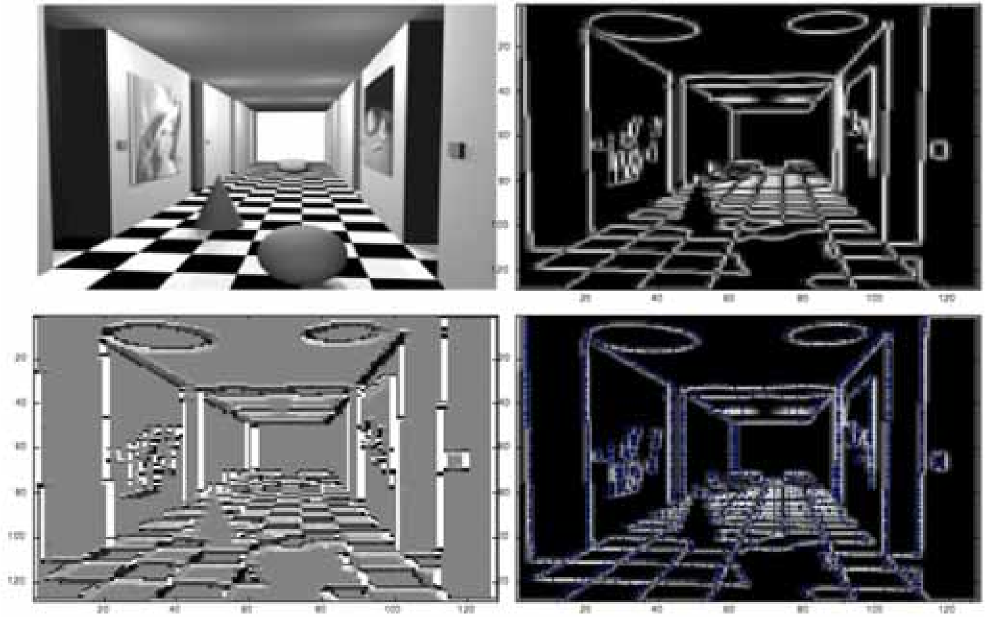

Fig. 3. Top Left: Original image, Top Right: Wavelet Transform Modulus, Bottom Left: wavelet transform modulus phase, Bottom Right: Wavelet Transform Modulus Maxima with Phase vectors

Where

$$\alpha(s,k) = tan^{-1}\left({D_{v,s,k}}\big/{D_{h,s,k}} \right) \tag{11}$$

The vector $\vec{n}(k)$ points to the direction normal to the edge surface as

$$\vec{n}(s,k) = \left[cos\left(\Theta_{W,s,k}\right), sin\left(\Theta_{W,s,k}\right)\right] \qquad (12)$$

An edge point is the point p at some scale s such that $WT_{s,k}$ is locally maxima at $k = p$ and $k = p + \varepsilon\vec{n}(k)$ for ε small enough. These points are known as *wavelet transform modulus maxima* (*WTMM*), and are shift invariant through the wavelet transform. For further details in reference to wavelet modulus maxima and its translation invariance, reader is kindly referred to (S. Mallat 1999).

4. Correspondence estimation

The matching process of the proposed algorithm is categorized into two major parts. The first part of the algorithm defines the correspondence estimation process only at the coarsest scale level, whereas the second part defines the iterative matching process from finer up to the finest scale level. Correspondence estimation at the coarsest scale is the most important part of the proposed algorithm as the algorithm uses the hierarchical approach for correspondence estimation. Therefore, the part of the algorithm related to the correspondence estimation at finer scale levels is very much dependent on the outcomes of coarsest level matching. Finer level matching involves the local search at the locations where any predecessor candidates have already been selected, in the coarsest level. A block diagram, as shown in Figure 4, presents a detailed visual representation of the correspondence estimation algorithm.

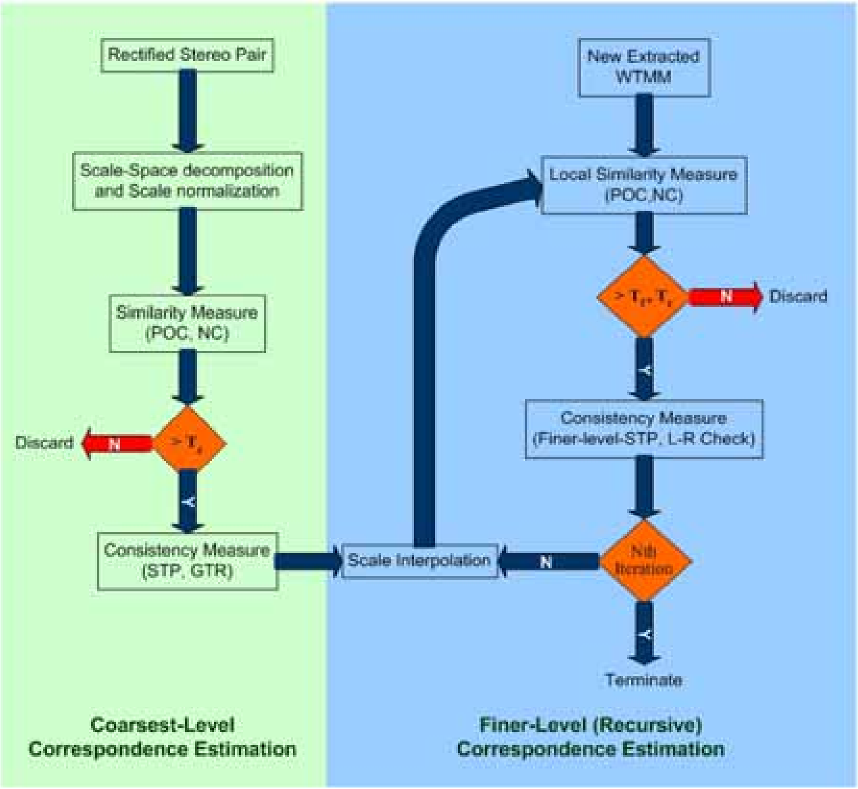
Rectified Stereo Pair
Scale-Space decomposition and Scale normalization
Similarity Measure (POC, NC)
Discard
N
>T
Y
Consistency Measure (STP, GTR)
Scale Interpolation
New Extracted WTMM
Local Similarity Measure (POC,NC)
>T, T
N
Discard
Y
Consistency Measure (Finer-level-STP, L-R Check)
Nth Iteration
N
Y
Terminate
Coarsest-Level Correspondence Estimation
Finer-Level (Recursive) Correspondence Estimation

Fig. 4. Block diagram of the correspondence estimation algorithm

4.1 Coarsest-level correspondence estimation

Coarsest level matching (*CLM*) is very important and crucial step of the whole matching process as correspondence estimation at finer levels are very much dependent on the outcome of *CLM*. All matching candidates at finer levels are arranged according to the matched locations found at the coarsest level. Considering the significance of (*CLM*) in the overall matching process there is a great need of keeping this process error free as much as possible. For this purpose, a comprehensive check is performed to exploit the likelihood of each candidate to be a credible match, before accepting or discarding it.

The matching process starts with wavelet decomposition up to level *N*, usually taken within the range of [4-5] depending on the size of the image. Before proceeding to the similarity measure block, wavelet scale normalization (*WSN*) is performed along with normalized correlation measure that helps to minimize the effect of illuminative variation that could exist in between the perspective views. The reason this comprehensiveness normalization is the nature of the application that we are trying to address in this work. The objects that we are concerned with, for the depth estimation, are aluminium die-castings with highly shiny and reflective surfaces. Therefore, there is a great need for the illuminative variation compensation before proceeding to the main correspondence estimation block.

The *WSN* is performed on each level of wavelet transform decomposition. It is done by dividing the details space with the approximation space and can be defined as

$$NW_{s,k} = {D_{dc,s,k}}\Big/{|A_{s,k}|} \qquad \forall\, dc \in \{h, v, d\} \tag{13}$$

Where {*h*,*v*,*d*} represents horizontal, vertical and diagonal detail components, respectively, *s* represents the scale of decomposition and A represents the approximation space.

4.1.1 Similarity measure

After the extraction of *wavelet transform modulus maxima* (*WTMM*) correlation based similarity measure is performed to obtain an initial estimation of the disparity map. A multiwindow approach (A. Fusiello 2000) is used to enhance the performance of correlation based similarity measure. The details about the improvements from single window to multiwindow approach can be found in (A. Fusiello 2000). The multi-window correlation score can be defined as

$$NC(s,k) = \overline{NC_{s,k,W_0} + \sum_{j=1}^{\frac{n_W}{2}} NC_{s,k,W_j}} \tag{14}$$

where NC_{s,k,W_0} represents the *normalized correlation* (*NC*) with respect to the central window whereas NC_{s,k,W_j} defines the *NC* respect to *j*th surrounding windows with n_W number of surrounding windows. In (14), the second term represents the summation of the best $n_w/2$ windows out of n_W. An average of the correlation scores from these windows is considered to keep the score normalized within the range of [*0-1*].

4.1.2 Probabilistic weighting

To make the correlation based selection criteria more comprehensive, a probabilistic weighting for the correlation measure in (14), is introduced. It is the probability of selection of a point, let say *i*th point, from within each search space, as a candidate C_i and out of r^2 search spaces as

$$P_c(C_i) = {}^{n_{C_i}}/_{r^2} \qquad where \qquad 1 \leq n_{C_i} \leq r^2 \tag{15}$$

where n_{C_i} is the number of times a candidate C_i is selected and r is the multiplicity of multifilter coefficients. As all matching candidates have equal probability of being selected, the probability of occurrence of any candidate through one search space is ${}^{1}/_{r^2}$. It is obvious from expression (15) that the $P_c(C_i)$ lies between the range of $\left[{}^{1}/_{r^2} - 1\right]$. We would like to call that probability term, *probability of occurrence* (POC) as it is the probability of any candidates C_i to appear n_{C_i} times in the selection out of r^2 search spaces. More specifically, if *j*th candidate C_j is selected ${}^{r^2}/_{2}$ times out of r^2 search spaces then POC of *j*th candidate is 0.5, i.e. $P_c(C_j)$=1/2. The correlation score in expression (15) is then weighted with POC as

$$CS_{C_i} = P_c(C_i) \sum_{n_{C_i}} NC_{C_i}(x,d) \qquad \forall_{n_{C_i}} \in \mathbb{Z}: n_{C_i} \leq r^2 \tag{16}$$

The probabilistic weighted correlation score CS_{C_i}, in (16), can be defined as *candidate strength* (*CS*). It represents the potential of the candidate to be considered for further processing and the involvement of the specific candidate in the selection of other potential candidates.

4.1.3 Symbolic tagging

Filtration of candidates, based on the *CS*, is followed by symbolic tagging procedure, which divides the candidates into three different pools based on three thresholds T_c, T_{c1} and T_{c2} possessing the criterion $T_{c2} > T_{c1} > T_c$. The threshold T_c acts as a rejection filter which filters out any candidate possessing lower *CS* than T_c. The rest of the candidates are divided into three pools as

$$\begin{aligned} NC_{s,k} &\geq T_{c_1}, && and && P_c(C_i) = 1, && \Rightarrow Op \\ NC_{s,k} &\geq T_{c_1}, && and && 0.5 \leq P_c(C_i) < 1, && \Rightarrow Cd \\ NC_{s,k} &\geq T_{c_2}, && and && {}^{2}/_{r^2} \leq P_c(C_i) < 0.5, && \Rightarrow Cr \end{aligned} \tag{17}$$

It can be seen from the first expression in (17), there is no ambiguity for the matches with tag *Op* as the *POC* is 1, whereas ambiguity does exist for the matches with tags *Cd* and *Cr*. Ambiguity is the phenomenon where there exists more than one correspondences for a single point in the reference image (R. Hartley 2003).

4.1.4 Geometric refinement

To address the issue of ambiguity, a simple geometric topological refinement is introduced in order to extract the optimal candidate matches out of the pool of ambiguous candidate matches. For that purpose, the geometric orientation of the ambiguous points with reference to Op from (17) is checked and the pairs having the closest geometric topology with respect to the Op are selected as optimal candidates. Three geometric features that are *relative distance difference* (*RDD*), *absolute distance difference* (*ADD*) and *relative slope difference* (*RSD*), are calculated to check the geometric orientation similarity. These geometric features are invariant through many geometric transformations, such as Projective, Affine, Metric and

Euclidean (M. Pollefeys 2000). The geometric measurement is then weighted with the *CS* of the candidates to keep the previous achievements of the candidates in consideration.

In order to calculate the geometric statistics a number of candidate pairs with tag *Op* are randomly selected. Let say n_r is the number of randomly selected pairs from n_{Op} candidate pairs possessing tag *Op*. Before proceeding to the calculation of *ADD* between the ambiguous pair of points we calculate average absolute distance *AAD* between selected pairs as

$$d_{Op_{n,i}} = \|Op_{1_i} - Op_{2_i}\|_{n_r\ :\ n_r \leq n_{Op}} \tag{18}$$

Where ||–|| defines the Euclidean distance between the pair of points with tags *Op* referring to image 1 and 2, respectively. The process in (18) is repeated *n* times to obtain *n* values of *AAD* in order to minimize the involvement of any wrong candidate pair that could have assigned the tag *Op*. Similarly for ambiguous candidate pairs with tag *Cd* the absolute distance can be calculated as

$$d_{Cd_j} = \left\|C_{Cd,1} - C_{Cd,2_j}\right\|_{m:\ j = 1 \cdots m} \tag{19}$$

Where m is the number of candidates $C_{Cd,2_i}$ selected from second image with potential to make a pair with $C_{Cd,1}$ in the first image. From (18) and (19) we can define ADD as

$$d_{A_{C_i}} = \left|\overline{\frac{d_{Cd_i} - d_{Op_{n,i}}}{d_{Cd_i} + d_{Op_{n,i}}}}\right|_n \tag{20}$$

Where $d_{A_{C_i}}$ is the *ADD* for *i*th candidate in the second image related to $C_{Cd,1}$ in the first image. Obviously we are interested in the candidate with minimum *ADD*. It is worth mentioning that absolute distances are invariant through Euclidean Transformation (R. Hartley 2003).

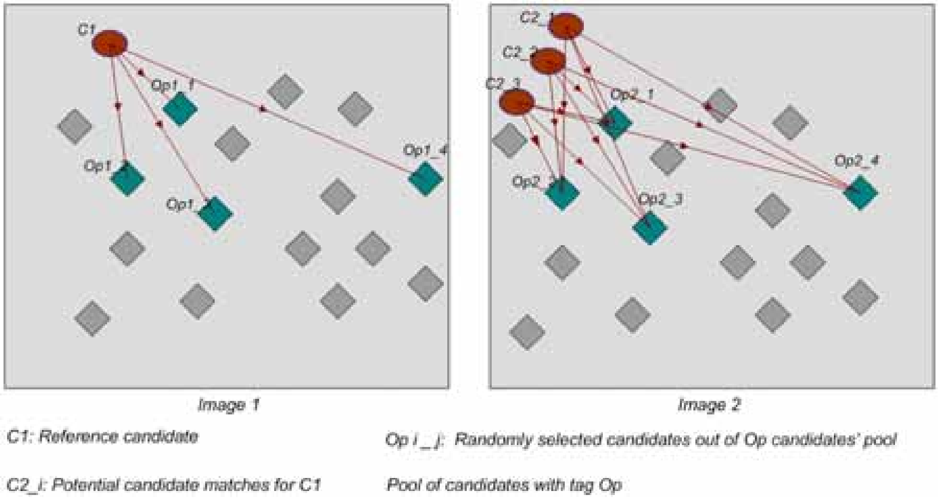
C1
Op1_1
Op1_2
Op1_3
Op1_4
C2_1
C2_2
C2_3
Op2_1
Op2_2
Op2_3
Op2_4
Image 1
Image 2
C1: Reference candidate
Op i _j: Randomly selected candidates out of Op candidates' pool
C2_i: Potential candidate matches for C1
Pool of candidates with tag Op

Fig. 5. Geometric refinement procedure

Before proceeding to the definition of *RDD* it is worthwhile to visualize the geometri refinement procedure as in Figure 5. There the candidate C_1 in the first image pairs with three potential candidates C_{2i} in the second image. The pairs with tag *Op*, shown by the gray colour, are spread all over the image and act as reference points in addressing the problem of ambiguity. The points with green colour are randomly selected points out of the pool of reference points with tag *Op*. Similarly, *RDD* can be defined by the following expression

$$d_{R_{C_i}} = min_j \left(\frac{d_{R_{C_{1,i}}} - d_{R_{C_{2,i,j}}}}{d_{R_{C_{1,i}}} + d_{R_{C_{2,i,j}}}} \right)_n \tag{21}$$

Where

$$d_{R_{C_{1,i}}} = \left\| C_1 - Op_{1,i} \right\|_{i \in n_{Op}} \quad where \quad i = 1 \cdots n_r \tag{22}$$

And

$$d_{R_{C_{2,i,j}}} = \left\| C_{2,j} - Op_{2,i} \right\|_{i \in n_{Op}, j \in m} \quad where \quad i = 1 \cdots n_r,\ j = 1 \cdots m \tag{23}$$

Similar to *ADD*, *RDD* is also calculated *n* times to minimize the effect of any wrongly chosen point with *Op* tag. Finally to calculate the relative slope difference we need to define relative slope for both images and between candidate points and the reference points. Thus, *RSD* can be defined as

$$d_{S_{C_i}} = min_j \left(\overline{\left| \frac{S_{C_{1,i}} - S_{C_{2,i,j}}}{S_{C_{1,i}} + S_{C_{2,i,j}}} \right|_n} \right) \tag{24}$$

The term (–)$_n$ defines the average over *n* repetitions, where *n* is usually taken within the range of [*3-5*]. Using (20), (21) and (24) a general and common term, as a final measure, to select the optimal candidate out of m potential candidates, is defined. The final term is weighted with the correlation score of the candidates from (16) to make the geometric measure more comprehensive as

$$Gc_i = max_i \left(CS_{C_i} \left(\overline{e^{-d_{A_{C_i}}} + e^{-d_{R_{C_i}}} + e^{-d_{S_{C_i}}}} \right) \right) \tag{25}$$

The expression in (25) could be defined as *geometric refinement score* (*GRS*). The candidate with the maximum *GRS* is then selected as optimal match and will be promoted to the symbolic tag of *Op*.

4.1.4 Scale interpolation

The disparity from coarser disparity d_c to finer disparity d_F is updated according to

$$d_F = d_L + 2d_c \tag{26}$$

Where d_L is the local disparity obtained within the current scale level. This process is repeated until the finest resolution is achieved which is the resolution of input image. An example of the outcome of coarsest level correspondence estimation can be seen in Figure 6.

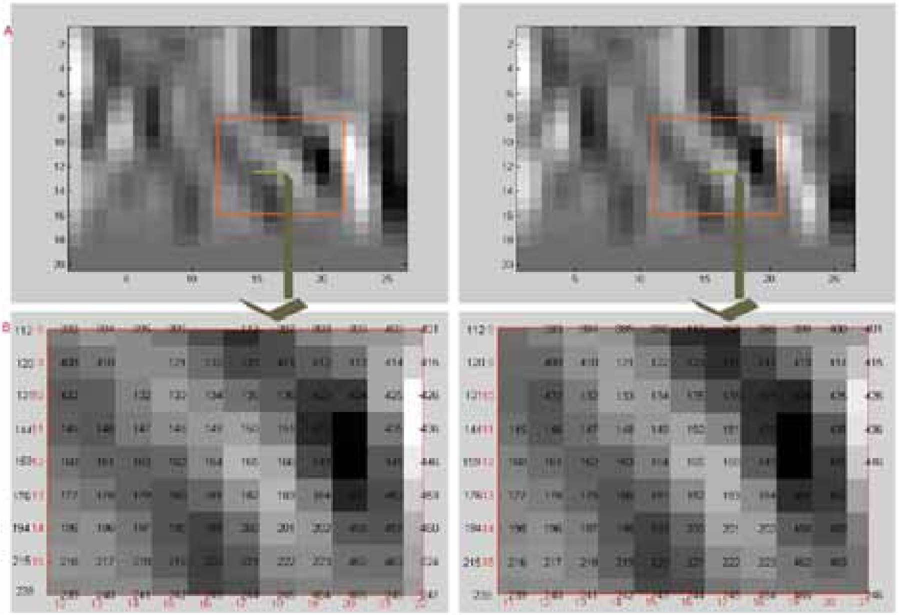

Fig. 6. An outcome of coarsest level matching

4.2 Finer-level correspondence estimation

Correspondence estimation at the finer level constitutes an iterative local search, based on the information extracted from the coarsest level. Due to the simpler nature of this search no geometric optimization is performed to deal with ambiguity but rather simpler approach is used, generally known as *left-right consistency* (*LRC*) check and can be defined as

$$-d_{s,k,1} = d_{s,k,2}\left((s,k)_x + d_{s,k,1}\left((s,k)_x,(s,k)_y\right),(s,k)_y\right) \qquad (27)$$

Where $d_{k,i}$ is the estimated discrete disparity of kth coefficient in ith image, whereas k_x and k_y are the x and y coordinates of the kth coefficient at scale s.

Similar to coarsest level search, interpolated coefficients are assigned candidate strength based on correlation scores and probability of occurrence as in (16), however the search is only defined over the relevant interpolated areas. Before proceeding further to the symbolic tagging procedure and *LRC* check, any coefficient with insignificant *CS* is discarded.

Symbolic tagging procedure is very similar to the one presented in section 4.1.3 however new assignation of tags depends on their ancestors' tags. In other words the coefficients that are interpolated from the coefficient, at the coarsest level, with tag *Op* will be dealt with different conditions than the one with tag *Cd*. The coefficients interpolated from *Op* are assigned tags as

$$\forall\, Op \Rightarrow \begin{cases} Op & if \quad P(C_k) \geq 0.5,\ \ NC(s,k) \geq T_{f_1} \\ Cd & if \quad P(C_k) \geq 0.2,\ \ NC(s,k) \geq T_{f_1} \end{cases} \qquad (28)$$

Whereas the coefficients having predecessor with tag *Cd*, we have

$$\forall\, Cd \Rightarrow \begin{cases} Op & if \quad P(C_k) = 1,\ NC(s,k) \geq T_{f_1} \\ Cd & if \quad P(C_k) \geq 0.2,\ NC(s,k) \geq T_{f_1} \end{cases} \tag{29}$$

Where T_{f_1} is usually chosen within the range of [0.4 - 0.5]. Similar to the coarsest level matching, the coefficients with Op are considered as the reference locations that will assist in rearranging the *Cd* coefficients using the expression in (28).

After this step, some gaps still left in the disparity map which is required to be filled to achieve dense depth map. These gaps are due to coefficients that were not taken into account before due to the unavailability of linked ancestors and have just appeared in the current scale. These coefficients are assigned *Cd* if and only if their strength i.e. *CS* from (16), is greater than T_{c1} and perform best in *LRC* check provided in (27).

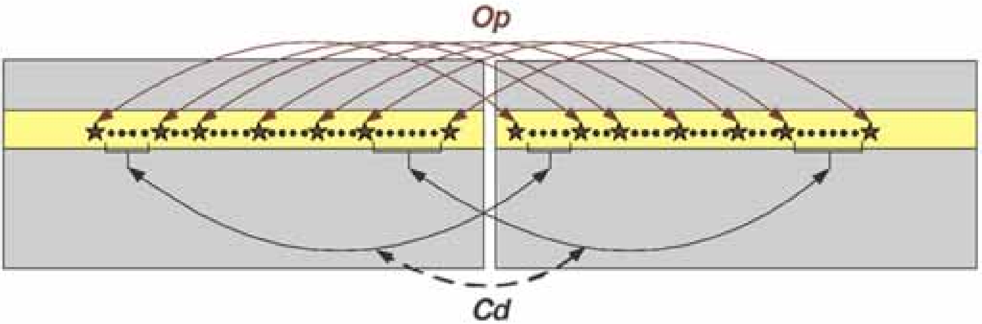

Fig. 7. Local search constellation relation between the images after symbolic tagging

The process of finer level correspondence estimation is repeated until the finest resolution is achieved, i.e. the resolution of the input image. Using a number of thresholds and symbols makes the appearance of the algorithm a little bit complicated and computationally expansive however comparing to existing algorithms it is not much different. Currently no explicit comparative information is extracted to support our claim but is intended for future works. No argument, there are many correlation based algorithms that are very fast due to low computational cost but dose not provide very promising qualitative performance. In addition, most of the algorithms, existing in the literature, perform post processing to coverup the deficiencies occurred during the correspondence estimation process which is itself is very computationally expansive. On the other hand proposed algorithm, due to the comprehensive criteria of selection/rejection, dose not require any post processing.

Furthermore, due to hierarchical nature, the disparity search is only $2^{level-1}$ of the original disparity search required at the input image level, that is, for a required search of 32 the proposed algorithm only required to search 4 disparities with decomposition of level 4.

5. Disparity estimation

The algorithm presented is exploited to its maximum capacity in terms of the stereo correspondence estimation performance. Four popular synthetic images are chosen from the database of the University of Middlebury. The relevant disparity maps are shown in Figure 8 to 9. In addition, error images are also calculated for each of the estimated disparity maps that simply are the absolute difference, defined in (30), between the ground truth and estimated disparity maps in terms of gray scale intensity values, as shown in subfigures 8(D) and 9(D). The absolute error can be expressed as

$$E = |d_G(x,y) - d_E(x,y)|_{\forall\, x \in X,\mathbb{Z},\ y \in Y,\mathbb{Z}} \quad (30)$$

Where $d_G(x,y)$ is the discrete ground truth disparity map, whereas $d_E(x,y)$ is the estimated one.

In order to find the statistical deviation of the estimated disparity maps from the provided ground truth disparity, two statistics are calculated as

$$\boldsymbol{R} = \sqrt{\frac{1}{N} \sum_{x,y} |\boldsymbol{d_E}(\boldsymbol{x},\boldsymbol{y}) - \boldsymbol{d_G}(\boldsymbol{x},\boldsymbol{y})|^2} \quad (31)$$

and

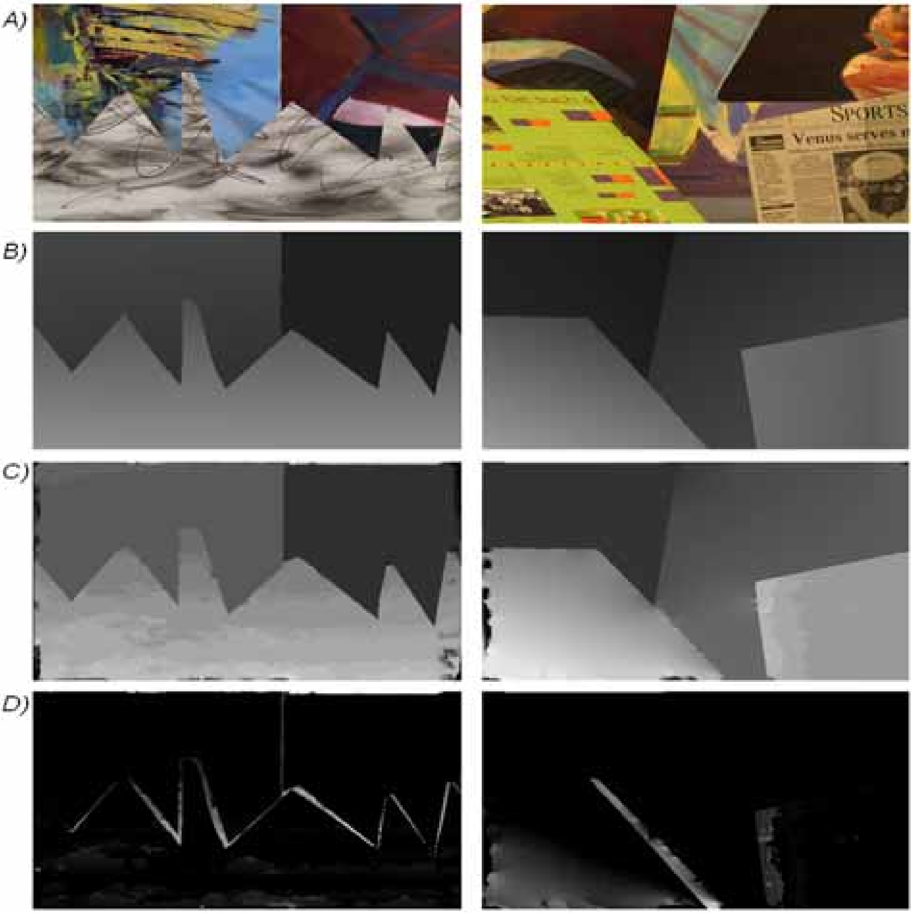

Fig. 8. A) Right images of the Sawtooth (left) and Venus (right) stereo pair, B) Ground truth disparity maps, C) Estimated disparity maps, D) Disparity error

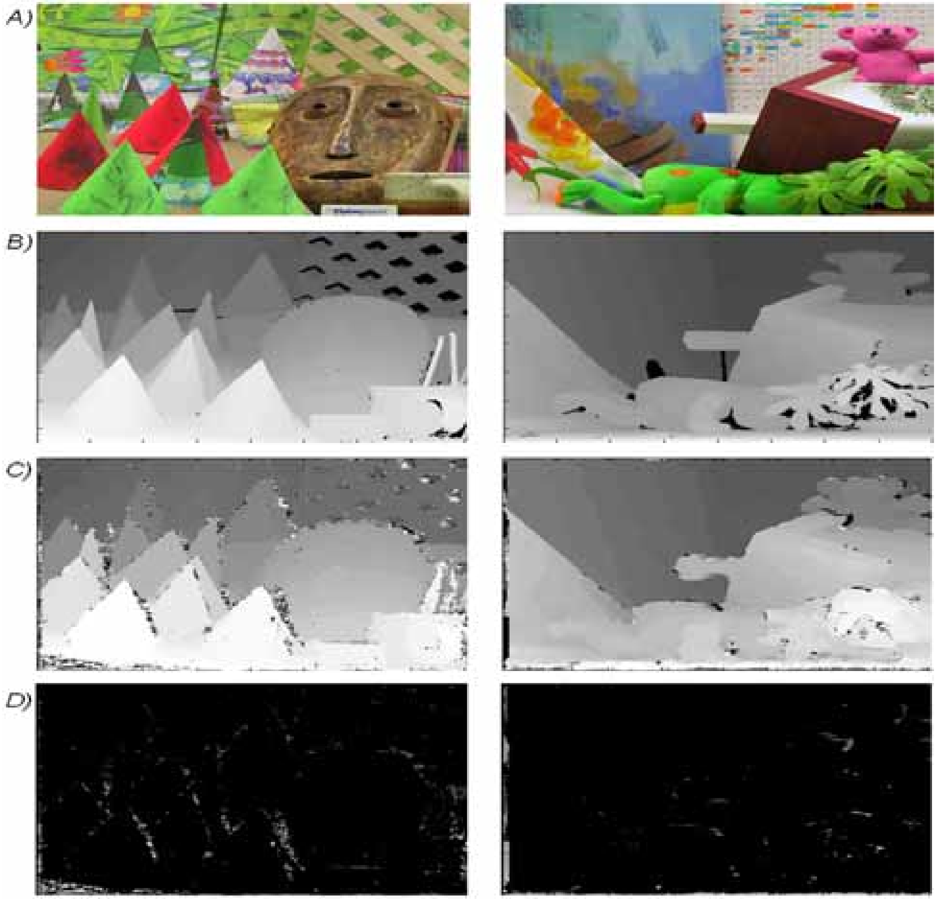

Fig. 9. A) Right images of the *Cones* (left) and *Teddy* (right) stereo pair, B) Ground truth disparity maps, C) Estimated disparity maps, D) Disparity error

$$\boldsymbol{B} = \frac{1}{N}\sum_{x,y} |\boldsymbol{d_E}(\boldsymbol{x},\boldsymbol{y}) - \boldsymbol{d_G}(\boldsymbol{x},\boldsymbol{y})|^2 > \boldsymbol{\xi} \tag{32}$$

Where *R* and *B* represent the *Root Mean Squared Error* (*RMSE*) and *Percentage of Bad Disparities* (*PBD*), respectively. *N* represents the total number of pixels in the input image whereas ξ represents the acceptable deviation of the estimated disparity value from the ground truth and is fixed to 1 in this particular work.

The images are taken into consideration with different complexities, in terms of pixel intensity variation and surface boundaries. First pair of stereo images is shown in Figure 8 with related ground truth disparity maps, estimated disparity maps and the error between the ground truth and estimated disparity maps. As it can be seen in Figure 8 the edges of the discontinuities are extracted to high accuracy and estimated disparity is very much similar to the ground truth disparity, visually. The *RMSE* and *PBD* score for *Sawtooth* and *Venus* are *R* = 1.9885, *B* = 0.0262 and *R* = 0.1099, *B* = 0.0381, respectively.

Similarly, another pair of disparity maps are shown for images *Cones* and *Teddy* and related *RMSE* and *PBD* scores are R = 3.3798, B = 0.1270 and R = 2.7629, B = 0.1115, respectively, as shown in Figure 9.

6. Disparity estimation

To further validate the claims about the performance of the proposed algorithm a comparison is performed between the proposed algorithm and a number of selected algorithms from the literature. Eight algorithms are chosen, known for their performance, within the computer vision research community. These estimated disparity maps are related to the images *Cones, Venus* and *Teddy*. The chosen algorithms for comparison purpose are *Double-bp (Q`ıngxiong Yang 2006)*, *Graph Cuts (D. Scharstein 2002)*, *Infection (G. Olague 2005)*, *Layered (L. Zitnick 2004)*, *Scanline Optimization (D. Scharstein 1998)*, *SSD min. Filter (D. Scharstein 2002)* and *Symmetric-Occlusion (J. Sun 2005)*.

The estimated disparity map selected for comparison against the aforementioned algorithms from the literature is generated using MW2 (Özkaramanli H. Bhatti A. and 2002). These calculated statistics, i.e. R and B, for the analysis of comparative performance with respect to the estimated results are shown in Table 1 and Figures 10 to 12. It is obvious from Table 1 and Figures 10 to 12, the proposed algorithm has performed best in the case of *Venus* image. However, in case of *Cones* and *Teddy* images the proposed algorithm has ranked 3rd, though very competitive to the algorithm ranked 1. Specifically in case of B, the proposed algorithm has outperformed all other algorithms. This reflects the true consistency and robustness of the proposed algorithm as number of bad disparity values estimated are lowest in all cases. It also reflects the comprehensiveness of the selection criteria defined by consistency measure from expressions (16) and (25).

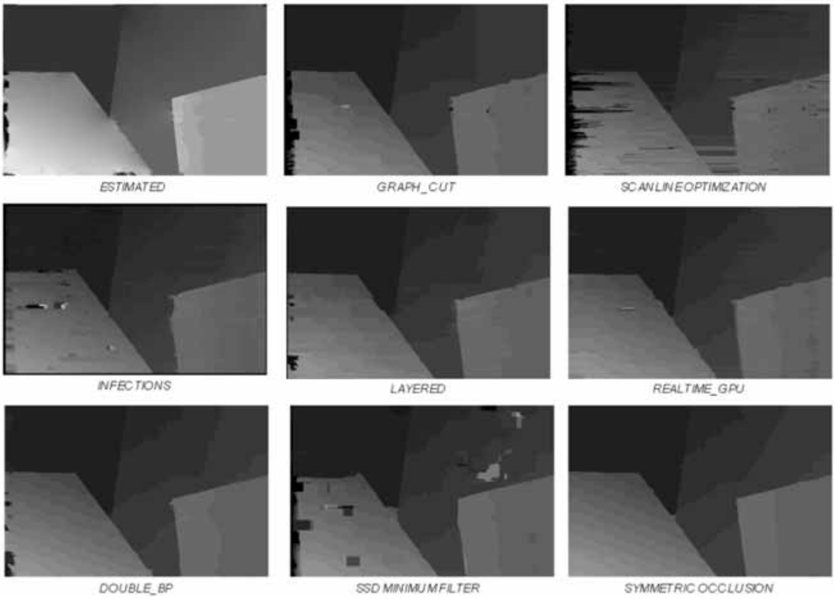

Fig. 10. Comparison of estimated disparity map with existing algorithms for image *Venus*

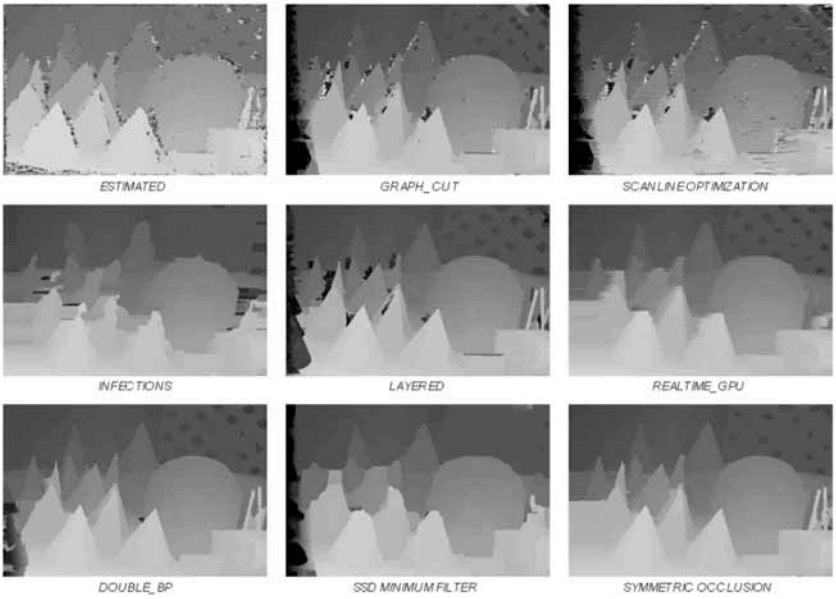

Fig. 11. Comparison of estimated disparity map with existing algorithms for image *Cones*

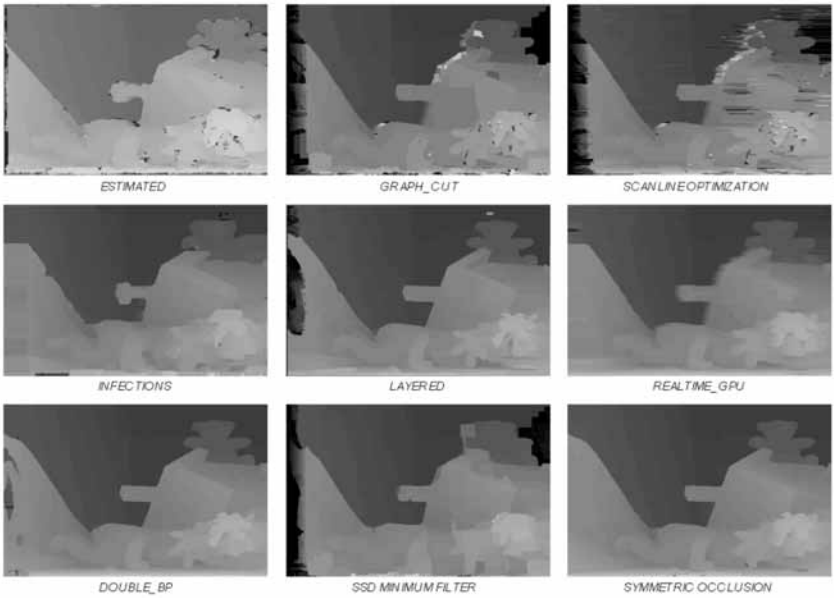

Fig. 12. Comparison of estimated disparity map with existing algorithms for image *Teddy*

Algorithms	Cones		Venus		Teddy	
	R	B	R	B	R	B
Estimated	3.3798_3	0.1270_1	1.9885_1	0.0262_1	2.7629_3	0.1115_1
Double-Bp	3.4898	0.2329	2.2114_3	0.2860_3	2.9360	0.2842
Graphcut	4.9694	0.2732	3.3977	0.3065	5.6912	0.3314
Infection	4.2949	0.2147_3	4.4952	0.3119	4.5092	0.2439_3
Layered	4.6167	0.2638	3.1955	0.3186	4.3622	0.3096
Realtime-Gpu	3.2784_2	0.2456	2.0780_2	0.2609_2	2.7535_2	0.2815
Scanline Opt.	5.4622	0.2989	4.2491	0.3090	5.7917	0.3538
SSD min. Filter	4.5599	0.2248	3.733	0.2960	5.9532	0.3111
Sym. Occlusion	3.1457_1	0.2145_2	15.7478	1.000	2.6445_1	0.2421_2

Table 1. A comparison of the estimated disparity with a number of existing well known algorithms

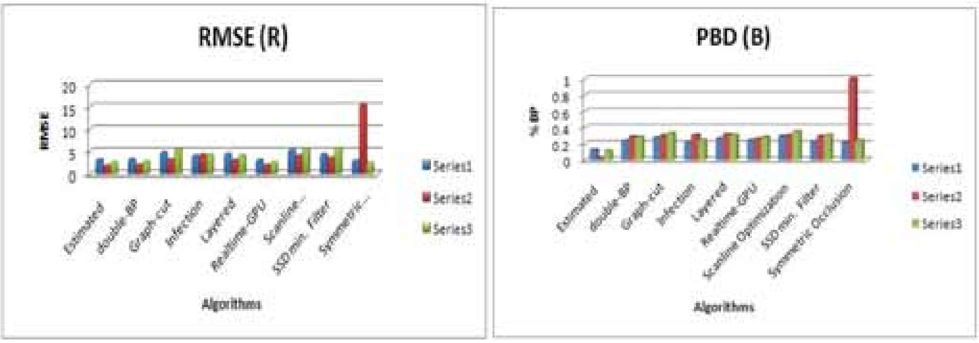

Fig. 13. comparison of the estimated disparity with a number of existing well known algorithms

7. Depth estimation

The main objective of the presented work is to obtain accurate depth estimations for quality inspection of the metallic parts in automotive industry. Consequently, it is important to keep the system simple with minimal hardware involvement, as the cost is an important factor for industry acceptance of the technology. While there are some existing techniques, such as (D. Scharstein 2003), capable of producing accurate disparity maps by using light or laser projectors however are both difficult and expensive to use in a typical production environment. In the presented work, only a basic hardware setup is used namely stereo cameras and circular florescent light. The simplicity of the hardware helps in keeping the implementation costs down and to make the automated fault detection in the automotive manufacturing industry as flexible as possible. Furthermore, hardware simplicity helps in deploying the system in different factory environments without a significant change in the setup.

For performance estimation of the proposed algorithm, two different metallic parts are used. In this section of the work, the main concern is to estimate accurate depth of the object, i.e. z dimension, and not the x, y as these dimensions can be estimated using a single 2D view. For the validation of the estimated depth map, one-dimensional cross-section of the depth difference between the estimated and the real depth maps are shown in Figures 14 and 15. One-dimensional cross-section of the depth difference is also shown for each of the two parts to give an idea about the quality of the estimated depth maps and their accuracy. As can be seen from the estimated depth maps the part defects can be accurately detected.

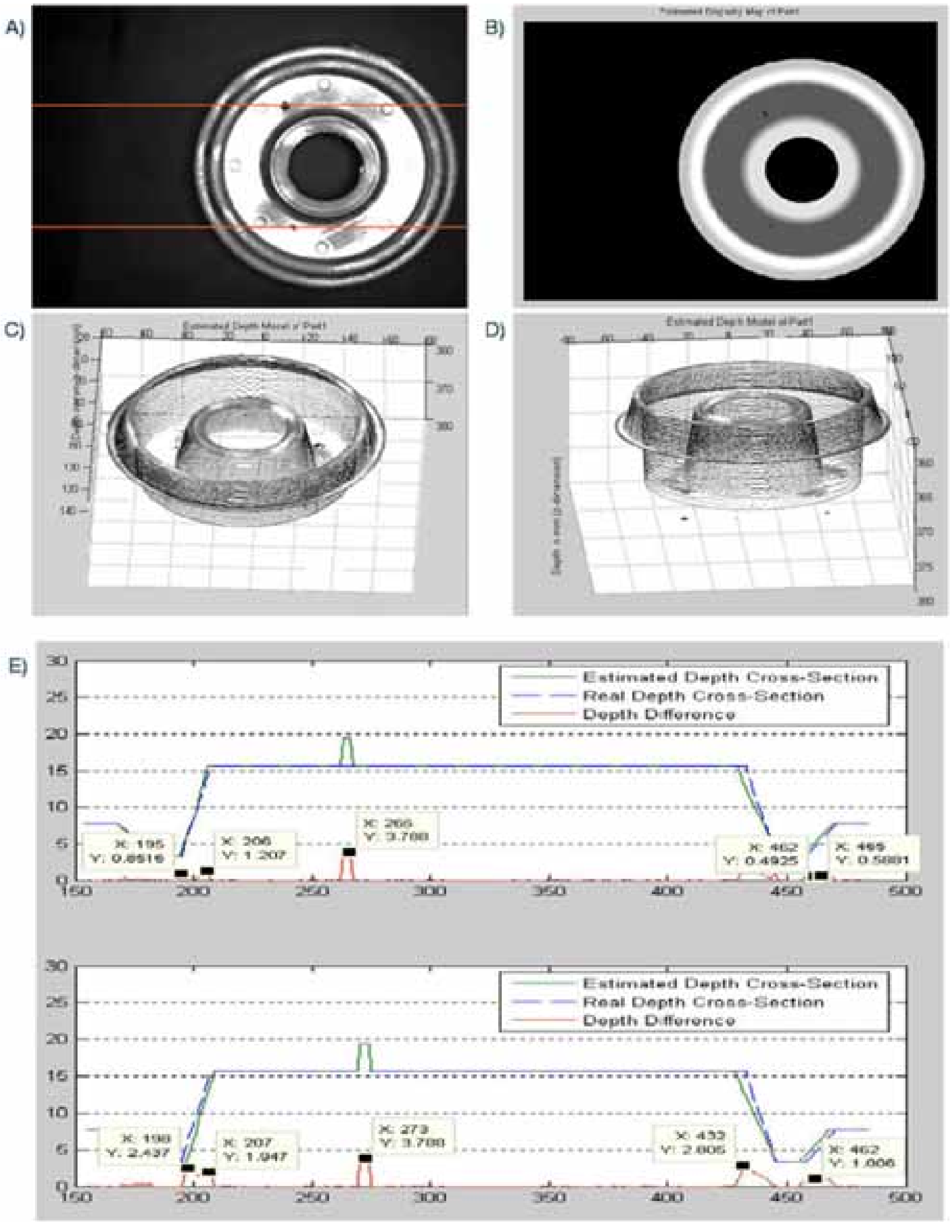

Fig. 14. A: Original image of Part-1 (Bad sample) B: Estimated Disparity Map of Part-1 C: Estimated 3D Depth of Part-1 in (mm) D: Estimated 3D Depth of Part-1 in (mm) (difference view) E: Different view of the Estimated Depth map (mm)

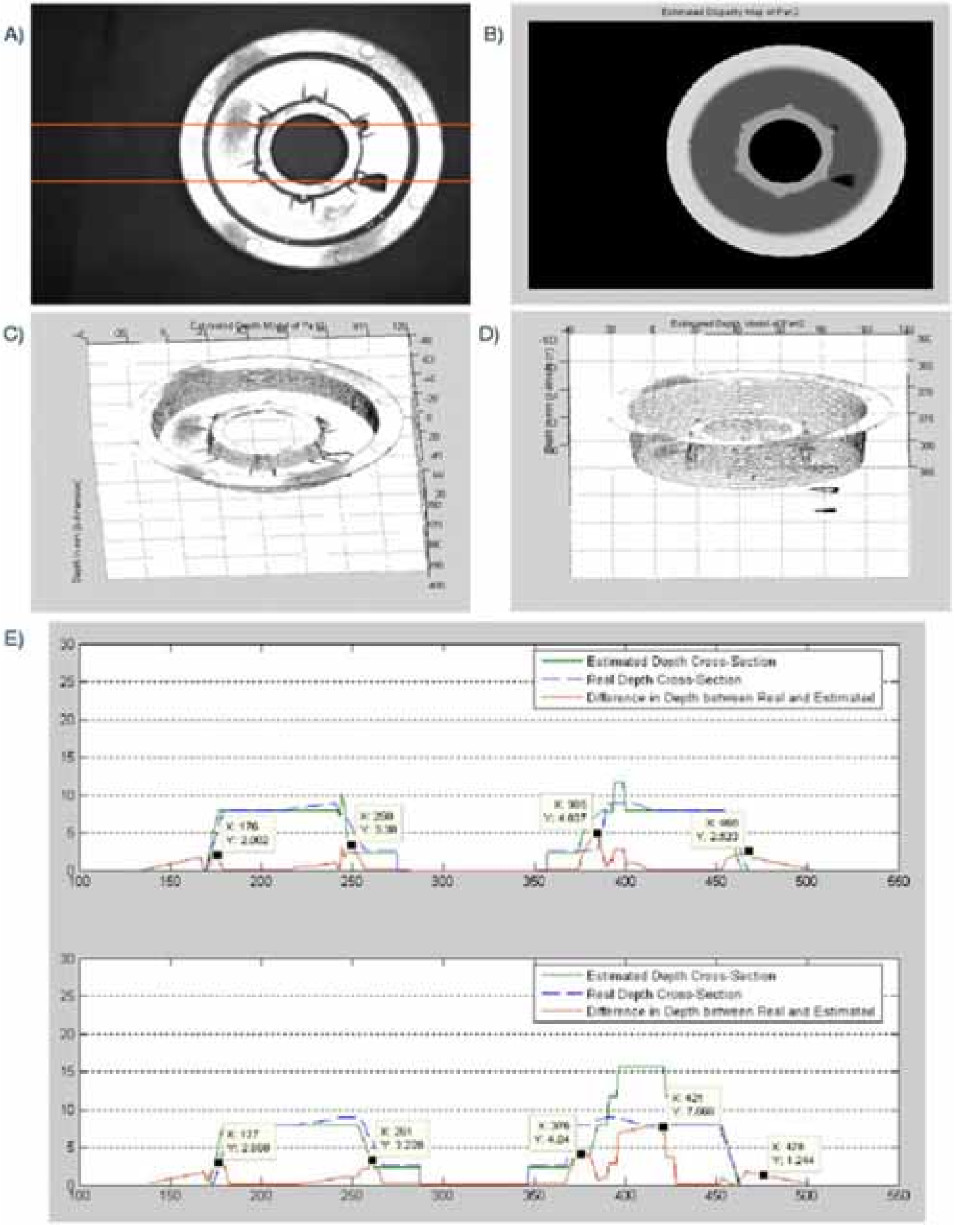

Fig. 15. A: Original image of Part-2 (Bad sample) B: Estimated Disparity Map of Part-2 C: Estimated 3D Depth of Part-2 in (mm) D: Estimated 3D Depth of Part-1 in (mm) (difference view) E: Different view of the Estimated Depth map (mm)

Referring to Figure 14 and 15, the difference between the real and estimated depth is very small across the area with no defect. For Part 1, the difference between the estimated and real depth lies within the range of [0.80-1.49*mm*] whereas for Part 2 the difference is [0.018-1.42 *mm*]. Therefore, the maximum depth difference regarding Part 1 and Part 2 is 1.49mm and 1.42mm, respectively. From the upper value of depth difference, an error tolerance can be set for differentiating between good and defective parts in an inspection system. Moreover, in Figure 14(E) and Figure 15(E) the sharp peaks are in fact due to the difference in x and y dimensions rather than the difference in depth.

8. Conclusion

The developed vision system consists of a novel robust algorithm. The proposed algorithm uses the stereo vision capabilities and multi-resolution analysis to estimate disparity maps and the concerned 3D depths. Furthermore, it uses multiwavelets theory that is a newer way of scale space representation of the signals and considered as fundamental as Fourier and a better alternative. The proposed algorithm uses the well-known technique of *coarse-to-fine* matching to address the problem of stereo correspondence. The translation invariant *wavelets transform modulus maxima* (*WTMM*) are used as corresponding features. To keep the whole correspondence estimation process consistent and resistant to errors, optimized selection criterion strength of the candidate is developed. The strength of the candidate involves the contribution of probabilistic weighted normalized correlation, symbolic tagging and geometric refinement. Probabilistic weighting involves the contribution of more than one search spaces, whereas symbolic tagging helps to keep the track of the most significant and consistent candidates throughout the process. Furthermore, geometric refinement addresses the problem of geometric distortion between the perspective views. The geometric features used in the geometric refinement procedure are carefully chosen to be invariant through many geometric transformations, such as affine, metric, Euclidean and projective. Moreover, beside that comprehensive selection criterion the whole correspondence estimation process is constrained to uniqueness, continuity and smoothness.

A novel and robust stereo vision system is developed that is capable of estimating 3D depths of objects to high accuracy. The maximum error deviation of the estimated depth along the surfaces is less than 0.5mm and along the discontinuities is less than 1.5mm. Similarly the time taken by the algorithm is with in the range of [12-15] seconds for the images of size [640-480]. The proposed system is very simple and consists of only a stereo cameras pair and a simple fluorescent light. The developed system is invariant to illuminative variations, and orientation, location and scaling of the objects, which makes the system highly robust. Due to its hardware simplicity and robustness, it can be implemented in different factory environments with out a significant change in the setup of the system. Due to its accurate depth estimation any physical damage, regarding the object under consideration, can be detected which is a major contribution towards an automated quality inspection system.

9. References

A. Bhatti, H. Özkaramanli (2002). M-Band multiwavelets from spline super functions with approximation order. International Conference on Acoustics Speech and Signal Processing, (ICASSP 2002), IEEE. 4: 4169-4172.

A. Fusiello, V. R., and E. Trucco (2000). "Symmetric stereo with multiple windowing." International Journal of Pattern Recognition and Artificial Intelligence.

A. Haar (1910). "Zur Theorie der orthogonalen Funktionen-Systeme." Math 69: 331-371.

Asim Bhatti, and Saeid Nahavandi, (2008). "Depth estimation using multiwavelet analysis based stereo vision approach." International Journal of Wavelets, Multiresolution and Information Processing 6(3): 481-497.

D. Scharstein, a. R. S. (1998). "Stereo matching with nonlinear diffusion." Int. J. of Computer Vision 28(2): 155-174.

D. Scharstein, a. R. S. (2002). "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms." Int. J. of Computer Vision 47: 7-42.

D. Scharstein, R. S. (2003). High-accuracy stereo depth maps using structured light. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). 1.

Fangmin Shi, Neil Rothwell Hughes, and Geoff Robert (2001). SSD Matching Using Shift-Invariant Wavelet Transform. British Machine Vision Conference: 113-122.

G. Olague, F. Fernández, C. Pérez, and E. Lutton, (2005). "The infection algorithm: an artificial epidemic approach for dense stereo correspondence." Artificial Life.

G. Plonka, and V. Strela, (1998). From Wavelets to Multi-wavelets. 2nd Int. conf. on Math. methods for curves and surfaces, Lillehammer, Norway: 375-400.

He-Ping Pan (1996). "General Stereo Matching Using Symmetric Complex Wavelets." Wavelet Applications in Signal and Image Processing 2825.

He-Ping Pan (1996). Uniform full-information image matching using complex conjugate wavelet pyramids. XVIII ISPRS Congress, International Archives of Photogrammetry and Remote Sensing. 31.

I. Cohen, S. Raz, and D. Malah, (1998). 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.

J. C. Olive, J. Deubler and C. Boulin, (1994). Automatic registration of images by a waveletbased multiresolution approach. SPIE. 2569: 234-244.

J. Magarey, and N.G. Kingsbury (1998). "Motion estimation using a complex-valued wavelet transform." IEEE Transections on signal proceessings 46(4): 1069-1084.

J. Margary, and A. dick (1998). Multiresolution stereo image matching using complex wavelets. Proc. 14th Int. Conf. on Pattern Recognition (ICPR). 1: 4-7.

J. Sun, Y. Li, S.B. Kang, and H.-Y. Shum, (2005). Symmetric stereo matching for occlusion handling. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE.

L. Di Stefano, M. M., S. Mattoccia, G. Neri (2004). "A Fast Area-Based Stereo Matching Algorithm." Image and Vision Computing 22(12): 938-1005.

L. Zitnick, S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, (2004). High-quality video view interpolation using a layered representation. SIGGRAPH.

M. Lin, and C. Tomasi, (2003). Surfaces with occlusions from layered stereo. CVPR: 710-717.

M. Pollefeys (2000). 3D modelling from images. Conjunction with ECCV2000. Dublin, Ireland.

M. Unser, and A. Aldroubi, (1993). "A multiresolution image registration procedure using spline pyramids." SPIE Mathematical Imaging 2034: 160-170.

O. Faugeras, Q. T. L., T. Papadopoulo, (2001). The Geometry of Multiple Images, MIT Press.

Özkaramanli H., Bhatti A. and Bilgehan B. (2001). Multiwavelets From Spline super functions with approximation order. International Symposium on Circuits and Systems, (ISCAS 2001), IEEE. 2: 525 - 528.

Özkaramanli H. Bhatti A. and, Bilgehan B., (2002). "Multi wavelets from B-Spline Super Functions with Approximation Order." Signal Processing, Elsevier Science: 1029-1046.

Q`ıngxiong Yang, L. W., Ruigang Yang, Henrik Stewenius and David Nister (2006). Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling. CVPR. 2: 2347-2354.

R. Hartley, and A. Zisserman (2003). Multiple View Geometry. Cambridge, UK, Cambridge University Press.

R. R. Coifman, a. D. L. D. (1995). Translation-invariant de-noising. Wavelet and Statistics, Lecture Notes in Statistics, Springer-Verlag: 125-150.

S. Mallat (1991). "Zero-Crossings of a Wavelet Transform,." IEEE Transactions on Information Theory 37: 1019-1033.

S. Mallat (1999). A Wavelet Tour of Signal Processing, Academic Press.

S. Mallat, a. S. Z. (1993). "Matching Pursuits With Time-Frequency Dictionaries." IEEE Transactions on Signal Processing 41(12): 3397-3415.

X. Zhou, and E. Dorrer, (1994). "Automatic image-matching algorithm based on wavelet decomposition." IAPRS 30(1): 951-960.