Current Development in Theory and Applications of Wavelets
Volume… Issue… 2008, Pages…
This paper is available online at http://www.pphmj.com

DEPTH ESTIMATION OF METALLIC OBJECTS USING MULTIWAVELETS SCALE-SPACE REPRESENTATION

A. BHATTI and S. NAHAVANDI

Intelligent Systems Research Lab.
Deakin University
Vic 3217, Australia

Abstract

The problem of dimensional defects in aluminum die-castings is widespread throughout the foundry industry and their detection is of paramount importance in maintaining product quality. Due to the unpredictable factory environment and metallic with highly reflective nature, it is extremely hard to estimate true dimensionality of these metallic parts, autonomously. Some existing vision systems are capable of estimating depth to high accuracy, however are very much hardware dependent, involving the use of light and laser pattern projectors, integrated into vision systems or laser scanners. However, due to the reflective nature of these metallic parts and variable factory environments, the aforementioned vision systems tend to exhibit unpromising performance. Moreover, hardware dependency makes these systems cumbersome and costly. In this work, we propose a novel robust 3D reconstruction algorithm capable of reconstructing dimensionally accurate 3D depth models of the aluminum die-castings. The developed system is very simple and cost effective as it consists of only a pair of

2000 Mathematics Subject Classification:

Keywords and phrases: 3D depth estimation, aluminum die-casting, inspection system, multiwavelets transform modulus maxima, disparity estimation, stereo vision, scale-space representation, geometric refinement.

The work was supported by the Cooperative Research Center for Cast Metals Manufacturing (CAST).

Received July 11, 2008

stereo cameras and a defused fluorescent light. The proposed vision system is capable of estimating surface depths within the accuracy of 0.5mm. In addition, the system is invariant to illuminative variations as well as orientation and location of the objects on the input image space, making 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. The proposed system is a major part of quality inspection system for the automotive manufacturing industry.

1. Introduction

The visual inspection in current manufacturing processes mainly depends on human inspectors whose performance is generally inadequate and variable. The accuracy of human visual inspection declines with monotonic behavior of the processes even though human visual system is very well adapted to the unpredictable environments. The visual inspection processes require observing the same type of images repeatedly to detect anomalies. Computer based visual inspection provides a viable alternative to human inspectors. Currently, most of the vision systems, specifically linked to industrial assembly and inspection, rely on 2D information, whereas the contribution of vision systems with 3D reconstruction capabilities is negligible. Some systems are available [14] capable of estimating depth to high accuracy, but are very much hardware dependent. These systems use light and laser pattern projectors integrated with vision systems or laser scanners. However, due to the reflective nature of the aluminum die-castings and unpredictable factory environments, these systems do not demonstrate promising performance. Furthermore, hardware dependency makes these systems cumbersome and costly. The presented work uses a very basic hardware setup consisting of a pair of stereo cameras and a simple defused florescent light. This makes the automated fault detection, in the automotive manufacturing industry, highly flexible to different factory environments without a significant change in the setup. The presented system does not require any predefined orientation and location during image capture.

In addition to hardware simplicity a robust multiresolution analysis based stereo vision algorithm is presented which uses the concept of multiwavelets scale space representation and coarse to fine hierarchical correspondence estimation. Finding correct corresponding points from more than one perspective views in the 3D reconstruction process using stereo vision is subject to a number of potential problems such as occlusion, ambiguity, illuminative variations, and radial distortions. A number of algorithms have been proposed to address some of the problems in stereo vision. The majority of them can be categorized into three broad classes, i.e., local algorithms (LA) ([7], [10]), global algorithms (GA) ([4], [15]) and hierarchical algorithms (HA). The algorithms belonging to the LA group try to establish the correspondences over locally defined regions in the image space. Correlations based techniques are commonly used to estimate the similarities based on image pixel intensities. Generally, LA perform well in the presence of rich textured areas but have tendency of relatively lower performance in the featureless regions. Furthermore, local search using correlation windows usually lead to poor performance across the boundaries of image regions. On the other hand, algorithms belonging to GA group deal with the correspondence as a global cost-function optimization problem. These algorithms usually do not perform local search but rather try to find a correspondence assignment that minimizes a global cost function. There is a clear consensus in the computer vision community that algorithms belonging to GA group has overall better performance over LA. However, these algorithms are not free of shortcomings. GA are dependent on how well the cost function represents the relationship between the disparity and some of its properties such as smoothness and regularity. Moreover, how close that cost function representation is to the real world. Furthermore, the smoothness parameters make the disparity map smooth everywhere which may lead to poor performance at image discontinuities. Another disadvantage of these algorithms is their computational complexity, which makes them unsuitable for real-time applications. HA also known as *coarse to fine*, on the other hand, do not explicitly state a global function that is to be minimized like GA but exhibits a behavior similar to iterative optimization algorithms, such as

([1], [20]), and generally operate on an image pyramid where results from coarser levels are used to constrain more local search at finer levels. The proposed algorithm falls into this last category of algorithms, i.e., HA. The proposed algorithm uses multiwavelets, one of the most famous and widely used scale space representation, for coarse to fine iterative search. Multiwavelet scale space representation and its effect and applications within the context of signal/image processing, is out of the scope of this work and readers are kindly referred to ([3], [11]).

The proposed algorithm is capable of estimating depth of aluminum die-castings to very high accuracy. The qualitative performance of the developed vision system relies on a novel disparity estimation algorithm, which involves the use of multiresolution analysis and scale-space representation using multiwavelets theory. The proposed algorithm could be considered as a middle approach possessing the features of both LA and GA due to its hierarchical nature [16]. The proposed algorithm uses well-known technique of coarse-to-fine matching to address the problem of stereo correspondence estimation using the multiwavelets transform modulus maxima (MWTMM) as corresponding features. A new comprehensive selection criterion called *strength of the candidate* (SC) is introduced unlike most of the existing algorithms where selection is solely based on different aggregation costs ([5], [18]) within the context of multiresolution analysis. 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. Some earlier versions of the developed system can be found in [2].

2. Wavelet and Multiwavelets Fundamentals

Classical wavelet theory is based on the refinement equations as given below

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

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

where c_h and w_h represent the scaling and wavelet coefficients. 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 function and wavelet are involved. The latter case leads to the notion of multi-wavelets. A multi-wavelet basis is characterized by r scaling and r wavelet functions. Here r denotes the multiplicity in the vector setting with $r > 1$. Multi-scaling functions satisfy the matrix dilation equation as

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

Similarly for the multi-wavelets the matrix dilation equation can be expressed as

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

where

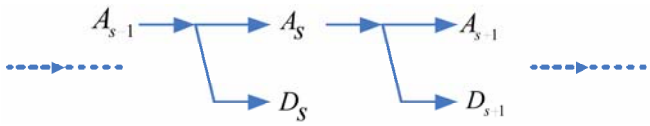

Figure 1. Wavelet theory based multiresolution analysis.

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

and

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

In equations (3) and (4), C_h and W_h are real $(r \times r)$ matrices of multi-filter coefficients whereas M represents the band of filter bank. Generally two band multiwavelets, i.e., $M = 2$, defining equal number of multi-wavelets as multi-scaling functions are used. For more information, about the generation and applications of multi-wavelets with, desired approximation order and orthogonality, interested readers are kindly referred to ([3], [11]).

A. Wavelet filter bank

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 represent 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 sub-spaces of A_{s-1}. Expression (7) can be better visualized by Figure 1.

The use of Mallat's dyadic filter-bank [11] results in three different detail space components, which are the horizontal, vertical and diagonal. Figure 2 can be the best visualize the graphical representation of the used filter-bank, where C and W represent the low-pass and high-pass filters consisting of the scaling functions and wavelets coefficients, respectively.

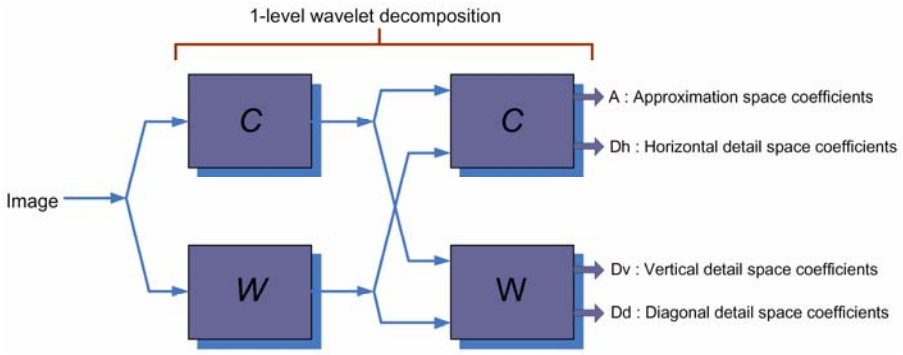

Figure 2. Mallat's dyadic wavelet filter bank.

B. 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 $W_{s,k}$ represents the magnitude of the WTM that can be expressed in terms of detail space coefficients as

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

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

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

where $\alpha(s,\ k)$ can be defined as

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

Based on $\Theta_{W_{s,k}}$ in (10), a vector $\vec{n}(s,\ k)$ can be defined that points normal to the edge surface as

$$\vec{n}(s,\ k) = [\cos(\Theta_{W_{s,k}}),\ \sin(\Theta_{W_{s,k}})] \tag{12}$$

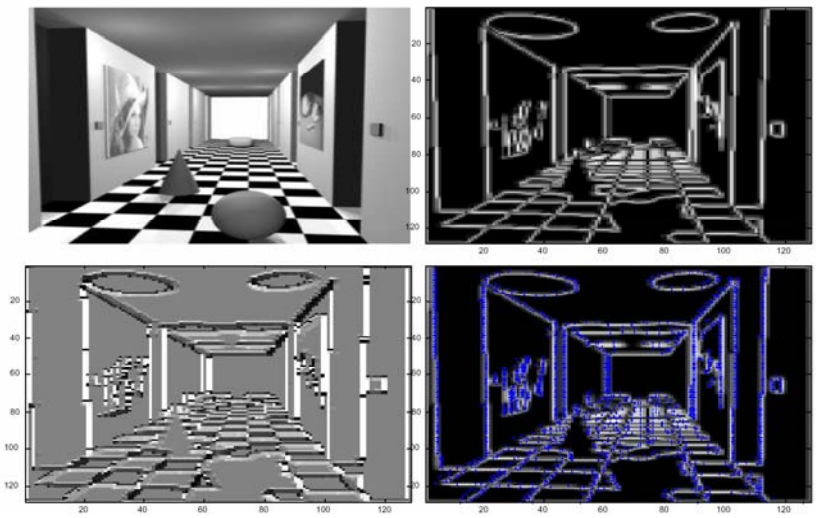

Figure 3. Top Left: Original image, Top Right: Wavelet Transform Modulus, Bottom Left: wavelet transform modulus phase, Bottom Right: Wavelet Transform Modulus Maxima with Phase vectors.

An edge point is the point p at some scale s such that $WT_{s,k}$ is locally maximum at $k = p$ and $k = p + \varepsilon\vec{n}(s, k)$ for $|\varepsilon|$ 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 [11].

3. Correspondence Estimation

The correspondence estimation process of the proposed algorithm is categorized into two major parts. First part of the algorithm defines the estimation only at the coarsest scale level, whereas the second part defines the iterative estimation process from finer up to the finest scale level. Correspondence estimation at the coarsest scale bears the utmost importance as the algorithm uses the hierarchical approach where correspondence estimations at finer scale levels are dependent on the outcomes of coarsest level processing. Therefore comprehensive selection criteria are introduced at this level for error free outcomes. Finer level correspondence estimation involves the local search at the locations

where correspondences have already been established during coarsest level processing.

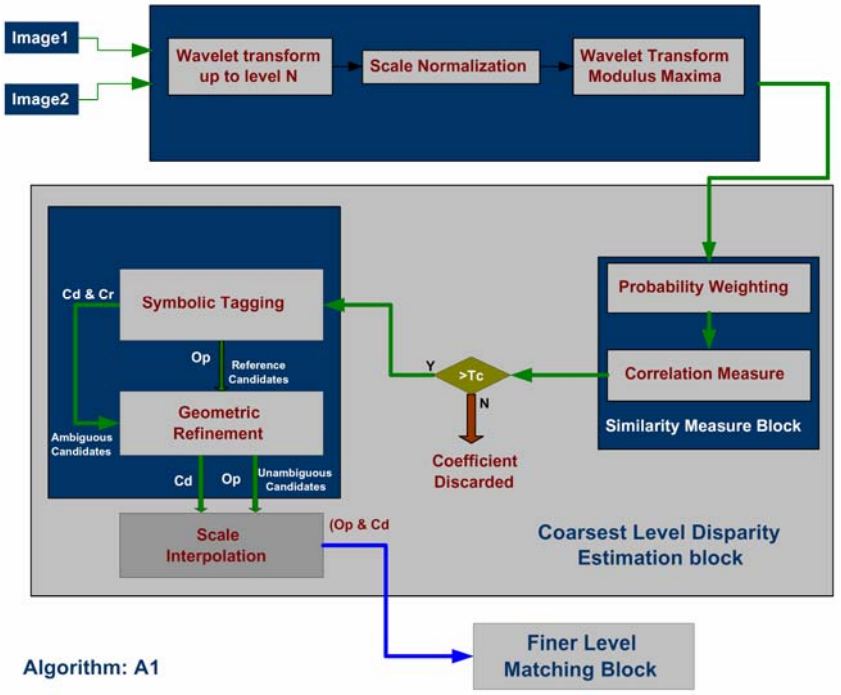

Figure 4. Block diagram of the correspondence estimation algorithm at the coarsest level.

A. Coarsest-level correspondence estimation

Coarsest level correspondence estimation (CLM) is very important and crucial step of the whole estimation process as finer levels estimation is dependent on the outcomes of CLM. All corresponding candidates at finer levels are arranged according to the correspondences established at the coarsest level. Considering the significance of CLM in the overall estimation process, there is a great need of keeping the estimation process error free to the highest possible extent. For this reason, a comprehensive selection criterion is introduced to exploit the likelihood of each correspondence to be credible, before accepting or discarding it. A block diagram, as shown in Figure 4, presents a detailed visual representation of the correspondence estimation algorithm at the coarsest level.

The correspondence estimation process starts with wavelet decomposition (transformation) up to level N, usually taken within the range of [4–15] depending on the size of the input images. Before proceeding to the similarity measure block as shown in Figure 4, wavelet scale normalization (WSN) is performed along with normalized correlation measure that helps in minimizing the effect of illuminative variations that could exist between the perspective views. The reason of this comprehensive normalization is the nature of the application that we are trying to address in this particular research work. The objects that we are interacting with are aluminum 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 transformation/decomposition. It is done by dividing the detail space components, i.e., $D_{h,v,d}$ with the approximation space component, i.e., A as in Figure 2 and can be defined as

$$NW_{s,k} = \frac{W_{s,k,dc}}{|A_{s,k}|}, \quad \forall dc \in \{h, v, d\}, \tag{13}$$

where h, v and d represent horizontal, vertical and diagonal detail components, respectively. Whereas s represents the scale of decomposition and k represents the kth coefficient.

(1) **Similarity measure.** After the extraction of wavelet transform modulus maxima (WTMM) correlation based similarity measure is performed to obtain an initial estimate of the disparity map ([9], [13]). A multi-window approach [8] is used to enhance the performance of correlation based similarity measure. General single window correlation expression can be expressed as

$$NC_{s,k} = \max_{X,Y}\left(\sum_{W} corr(NW^1_{s,k,W} \angle \Theta_{W_{s,k,W}}, NW^2_{s,k',W'} \angle \Theta_{W_{s,k',W'}})\right)_{\forall X,Y \in I}, \tag{14}$$

where $NW^i_{s,k,cw}$ represents kth scale normalized wavelet coefficient at

scale s surrounded by W and W' windows in reference to first and second images, respectively. Where I represents the WTM at any scale s. The multi-window correlation score using (14) can then be expressed as

$$NC(s,k) = NC_{s,k,W_0} + \sum_{j=1}^{n_W/2} NC_{s,k,W_j}, \tag{15}$$

where NC_{s,k,W_0} represents the correlation score with respect to the central window whereas NC_{s,k,W_j} represents the correlation score related to the surrounding windows. n_W represents the number of surrounding windows which is usually taken as 8, i.e., 9 window approach. The size of the windows used falls within the range of [9–13] coefficients. In (15), the second term represents the summation of the best $n_w/2$ windows out of n_W. An average of the correlation scores is considered to keep the score normalized within the range of [0–1].

(2) **Probabilistic weighting.** Just to refresh ones memory; wavelets/multiwavelets scale space decomposition produces r^2 components for each approximation and detail spaces at each level. However, in case of wavelet transformation $r = 1$, producing only one component for each approximation and detail space. Therefore the similarity measure, for each of the coefficients related to the reference image, is required to be performed throughout r^2 spaces to find the optimal correspondence. During this process, generally k'th coefficient gives the best similarity score for kth coefficient throughout r^2 spaces without reflecting any ambiguity. This usually happens in reference to prominent image features. However, there is a possibility that $\breve{k}$ or $\hat{k}$ can appear to be the better candidate than k' through some spaces r_p possessing the relationship $r_p < r^2$. One possible solution to overcome this aforementioned anomaly is to select the candidate with higher number of selections and discard the others. This approach is simpler but can cause serious consequences, especially if the number of hits of each selected coefficient is similar or very close to each other. To overcome the aforementioned issue a probabilistic weighting for the correlation

measure achieved through (15), is introduced. It is the probability of a feature, say k, to be the corresponding candidate C_k of any point k' throughout r^2 search spaces as

$$P(C_k) = n_{C_k}/r^2, \quad \text{where } 1 \leq n_{C_k} \leq r^2, \tag{16}$$

where n_{C_k} is the number of times a candidate C_k is selected and r is the multiplicity of multi-filter coefficients as given in (5) and (6). 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 (16) that the $P_c(C_k)$ lies between the range of $[1/r^2\ 1]$. This probability term is called *probability of occurrence* (**POC**) as it is the probability of any candidates C_k to appear n_{C_k} times in the selection out of r^2 search spaces. More specifically, if jth candidate C_j is selected $r^2/2$ times out of r^2 search spaces, then **POC**, i.e., $P(C_j) = 1/2$. The correlation score in expression (15) is then weighted with POC as

$$CS_{s,k} = P(C_k)\overline{\sum_{n_{C_k}} NC(s, k)}, \quad \forall_{n_{C_k} \in \mathbb{Z} : n_{C_k} \leq r^2}. \tag{17}$$

The probabilistic weighted correlation score $CS_{s,k}$, in (17), can be defined as *candidate strength* **CS** of kth coefficient. Where $\overline{\sum NC(s, k)}$ represents the average of n_{C_k} correlation scores. It represents the potential of the coefficient to be considered for further processing. In addition the candidates with $P(C_k) = 1$ assist in the selection of other potential coefficients.

(3) **Symbolic tagging.** Filtration of candidates, based on the **CS**, is followed by symbolic tagging procedure (**STP**), which divides the candidates into three different pools based on three thresholds T_c, T_{c_1} and T_{c_2} possessing the criterion $T_{c_2} > T_{c_1} > T_c$. 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

$$NC(s, k) \geq T_{c_1}, \text{ and } P(C_k) = 1, \Rightarrow \mathbf{Op},$$
$$NC(s, k) \geq T_{c_1}, \text{ and } 0.5 \leq P(C_k) < 1, \Rightarrow \mathbf{Cd},$$
$$NC(s, k) \geq T_{c_2}, \text{ and } 2/r^2 \leq P(C_k) < 0.5, \Rightarrow \mathbf{Cr}. \quad (18)$$

In general words the first expression in (18) represents that candidate C_k possesses probability 1, i.e., C_k is the only candidate selected through out r^2 search spaces for any reference point k with all $NC(s, k)_{\forall C_k \in r^2} \geq T_{c_1}$. It further defines that there is no ambiguity for the candidates **Op**. Second expression in (18) shows that candidate C_k is been selected through more than half, i.e., $r^2/2$ of the search spaces. This candidate is considered to possess potential of being the credible match, therefore is given a tag **Cd** with limited authority to go ahead. If this **Cd** tagged candidate fulfils the upcoming geometric criteria, presented in the next session, the tag will be changed to **Op** for this particular feature. Furthermore, the third expression defines the potential of the candidate as well, however in terms of very high correlation score rather than the **POC**. These candidates with tags **Cr** are automatically promoted to **Cd** after all the candidates with **Cd** are processed through geometric refinement.

Moreover, as the $P(C_k) \neq 1$, therefore ambiguity does exist for candidates with tags **Cd** and **Cr**. Ambiguity is the phenomenon where there exist more than one correspondence for a single point in the reference image [9]. The problem of ambiguity is addressed explicitly using a geometric optimization procedure presented in the next section. The rest of the candidates not fulfilling any of the above criteria are simply discarded, leaving only the most consistent correspondences.

(4) **Geometric refinement.** This section defines the geometric topological refinement (**GTR**) introduced to extract the optimal correspondences out of the pool of ambiguous correspondences. During this process, geometric orientation of the ambiguous candidates with respect to **Op** as in (18) is checked and the pairs having the closest

geometric topology 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 similarity of geometric orientation. The reason of selecting these geometric features, in order to address the problem of ambiguity, is their invariance through many geometric transformations, such as Projective, Affine, Matric and Euclidean [13]. The candidate strength in (17) is then weighted with the geometric measurement to keep the previous achievements of the candidates in consideration.

Before proceeding to the geometric refinement, it is worthwhile to visualize the geometric refinement procedure as shown in Figure 5. It can be seen that the candidate C_1 in the first image pairs with three potential candidates C_{2_i} in the second image. The geometric refinement procedure will assist in extracting the most optimal candidate. The pairs with tag **Op**, shown by the gray color, are spread all over the image and acted as reference points in addressing the problem of ambiguity. The points with green color are randomly selected points out of the pool of reference points with tag **Op**.

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

$$d_{Op_i} = \| \mathbf{Op}_{1_i} - \mathbf{Op}_{2_i} \|_{n_r : i=1 \cdots r,\ n_r \le n_{\mathbf{Op}}}, \tag{19}$$

where $\| \cdot \|$ defines the Euclidean distance between the pair of points with tags **Op** referring to images 1 and 2, respectively, where n_r is the number of randomly selected coefficients with tag **Op** out of n_{O_p} coefficients with tag **Op**. The common value of n_r is [7–10] using the assumption that at least n_r coefficients with tag **Op** are available. From (19) we can have

$$d_{Op} = \min_i (d_{Op_i}). \tag{20}$$

Similarly for ambiguous candidate pairs with tag **Cd** the absolute distance can be calculated as

$$d_{\mathbf{Cd}_j} = \| C_{\mathbf{Cd},1} - C_{\mathbf{Cd},2_j} \|_{m:j=1\cdots m}, \tag{21}$$

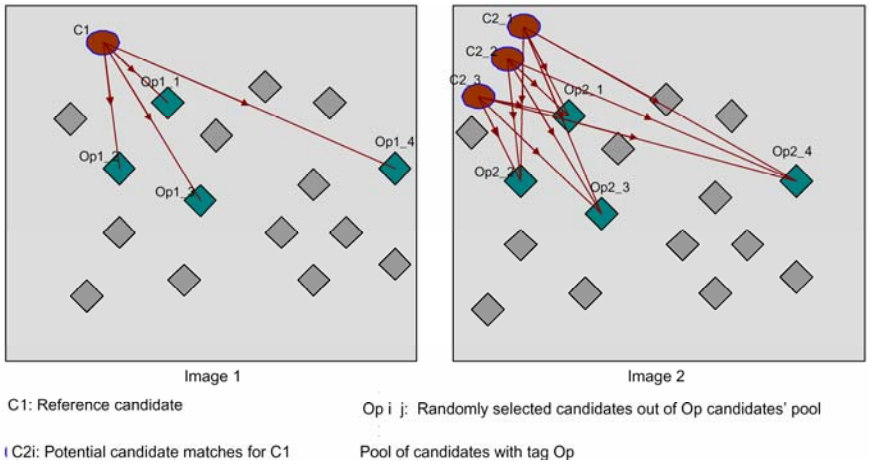

Figure 5. Geometric refinement procedure.

where m is the number of candidates $C_{\mathbf{Cd},2_j}$ selected from second image with potential to make a pair with $C_{\mathbf{Cd},1}$ in the first image. From (20) and (21) we can define the ADD as

$$d_{A_{C_j}} = \overline{\left(\left| \frac{d_{\mathbf{Cd}_j} - d_{Op}}{d_{\mathbf{Cd}_j} + d_{Op}} \right| \right)_n}, \tag{22}$$

where $d_{A_{C_j}}$ is the ADD for jth candidate in the second image related to $C_{\mathbf{Cd},1}$ in the first image. Obviously we are interested in the candidate with minimum ADD. The process in (22) is repeated n times to obtain an average value of AAD, over n repetitions, for each of the $d_{\mathbf{Cd}_j}$ candidate in order to minimize the involvement of any wrong candidate pair that could have been assigned the tag **Op**. Only the best value of AAD regarding any candidate j is considered. It is worth mentioning that absolute distances are invariant through Euclidean transformation [9]. Expression (22) assigns ADD values to all j candidates.

Similarly, RDD can be defined by the following expression:

$$d_{R_{C_j}} = \overline{\left(\min_i \left| \frac{d_{R_{C_{1,i}}} - d_{R_{C_{2,i,j}}}}{d_{R_{C_{1,i}}} + d_{R_{C_{2,i,j}}}} \right| \right)_n}, \tag{23}$$

where

$$d_{R_{C_{1,i}}} = \| C_1 - \mathbf{Op}_{1,i} \|_{i \in n_{\mathbf{Op}}}, \quad \text{where } i = 1 \cdots n_r \tag{24}$$

and

$$d_{R_{C_{2,i,j}}} = \| C_{2,j} - \mathbf{Op}_{2,i} \|_{i \in n_{\mathbf{Op}}, j \in m}, \quad \text{where } i = 1 \cdots n_r. \tag{25}$$

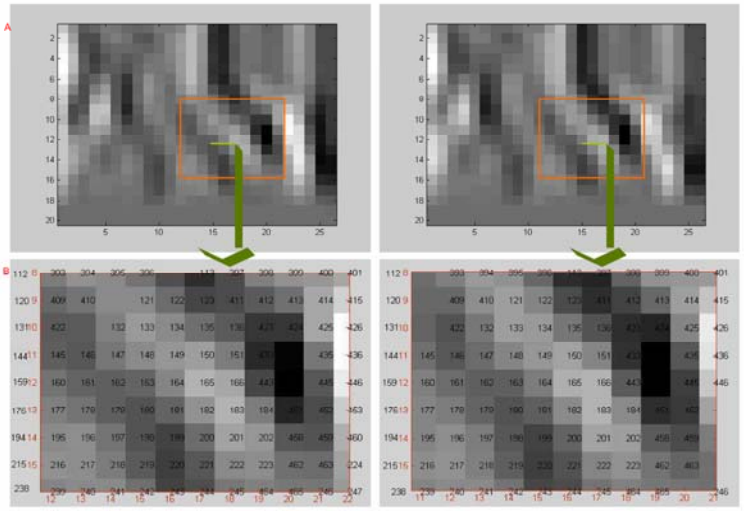

Figure 6. A: Stereo pair images at coarsest level B: Correspondence established within the cropped area

Similar to ADD, RDD is also calculated n times for each of the candidates j 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

$$d_{S_{C_j}} = \overline{\left(\min_i \left| \frac{S_{C_{1,i}} - S_{C_{2,i,j}}}{S_{C_{1,i}} + S_{C_{2,i,j}}} \right| \right)_n}. \tag{26}$$

The term $(.)_n$ defines the average over n repetitions, where n is usually taken within the range of [3–5]. Using (22), (23) and (26) 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 (17) to make the geometric measure more comprehensive as

$$Gc_j = \max_j (CS_{k=j} \overline{(e^{-d_{A_{C_j}}} + e^{-d_{R_{C_j}}} + e^{-d_{S_{C_j}}})}). \tag{27}$$

The expression in (27) could be defined as geometric refinement score (GRS). The candidate C_j with the maximum GRS is then selected as optimal match and will be promoted to the symbolic tag of **Op**.

An example of the coarsest level correspondence estimation is shown in Figure 6, where established coefficient correspondences are represented by same numbers on both sides. The image used is *Venus* taken from the website of Middlebury University and is decomposed to level 4 using *MW* multiwavelet basis from [12].

These correspondences are the outcome of the coarsest level correspondence estimation which then passes to finer level for further local search.

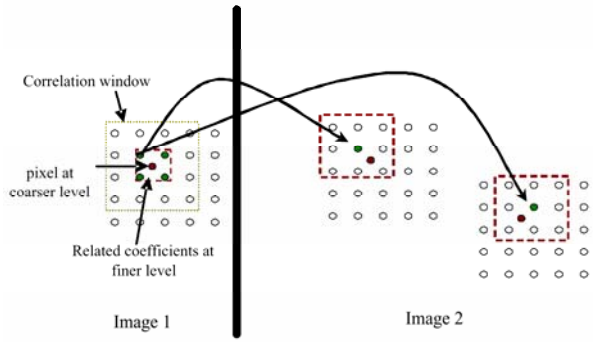

Figure 7. Scale Interpolation from coarser to finer level.

(5) **Scale interpolation.** The matching process at the coarsest level ends up with a number of established correspondences that are required 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. The interpolation process can be better visualized as in Figure 7, where green pixels represent the finer level pixels. 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.

As images are assumed to be rectified therefore correlation performed is omnidirectional, otherwise each location need to be checked for all 4 locations, represented by green. The matches are refined up to the finest level and leaving most consistent matches at the end of the process. The disparity from coarser disparity d_c to finer disparity d_F is updated according to

$$d_F = d_L + 2d_c, \tag{28}$$

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.

B. Finer-level correspondence estimation

Correspondence estimation at the finer level constitutes an iterative local search process, based on the information extracted from the coarsest level. Before highlighting the differences and features of the finer level correspondence estimation a block diagram is shown in Figure 8 for better visualization. Due to the simpler nature of this local search no geometric optimization is performed to deal with ambiguity but rather simpler approach of left-right consistency (LRC) check is used as

$$-d_{k,1} = d_{k,2}(k_x + d_{k,1}(k_x, k_y), k_y), \tag{29}$$

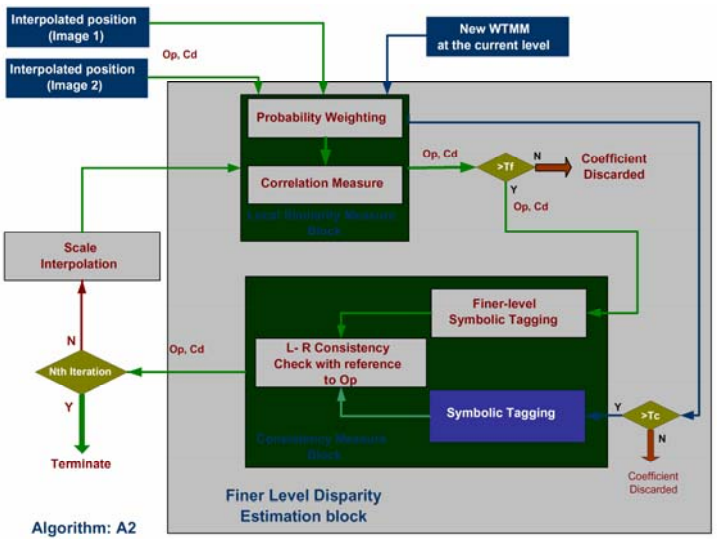

Figure 8. Block diagram of the correspondence estimation algorithm at finer levels.

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.

Similar to coarsest level search, interpolated coefficients are assigned candidate strength based on correlation scores as in (15) and probability of occurrence as in (17), 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, i.e., below certain threshold is discarded.

Symbolic tagging procedure is very similar to the one presented in Section 3.A.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 according to the following expression (30):

$$\forall \mathbf{Op} \Rightarrow \begin{cases} \mathbf{Op} & \text{if } P(C_k) \geq 0.5,\ NC(s,\ k) \geq T_{f_1} \\ \mathbf{Cd} & \text{if } P(C_k) \geq 0.2,\ NC(s,\ k) \geq T_{f_1}, \end{cases} \tag{30}$$

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

$$\forall \mathbf{Cd} \Rightarrow \begin{cases} \mathbf{Op} & \text{if } P(C_k) = 1,\ NC(s,\ k) \geq T_{f_1}, \\ \mathbf{Cd} & \text{if } P(C_k) \geq 0.2,\ NC(s,\ k) \geq T_{f_1}, \end{cases} \tag{31}$$

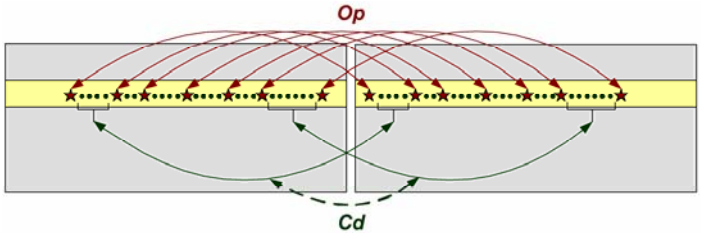

Figure 9. Local search constellation relation between the images after symbolic tagging.

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 (29).

After this step, some gaps still left in the disparity map which is required to be filled in order 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 (17), is greater than T_{c_1} and perform best in LRC check provided in (29).

The process of finer level correspondence estimation as shown in Figure 8 is repeated until the finest resolution is achieved, i.e., the resolution of the input image.

Using a number of thresholds and symbols make the appearance of the algorithm a little bit complicated and computationally expansive. Currently no explicit comparative information is extracted to support our claim but is intended for future works. There are many correlation based algorithms that are very fast due to low computational cost but does not provide very promising qualitative performance. In addition, most of the algorithms, existing in the literature, perform post processing to

cover up the deficiencies occurred during the correspondence estimation process which is itself very computationally expansive. On the other hand proposed algorithm, due to the comprehensive criteria for selection/ rejection, does not require any post processing. Moreover, due to hierarchical nature, the disparity search is only $2^{\text{level}-1}$ th 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. An example of the algorithmic outcome is shown in Figure 10.

4. Disparity Estimation

Before we proceed to the depth estimation process, which is the main objective of the presented work, a few raw disparity estimation results are presented. It is to provide an insight to the reader about the performance of the algorithm without any prior knowledge of the input images and

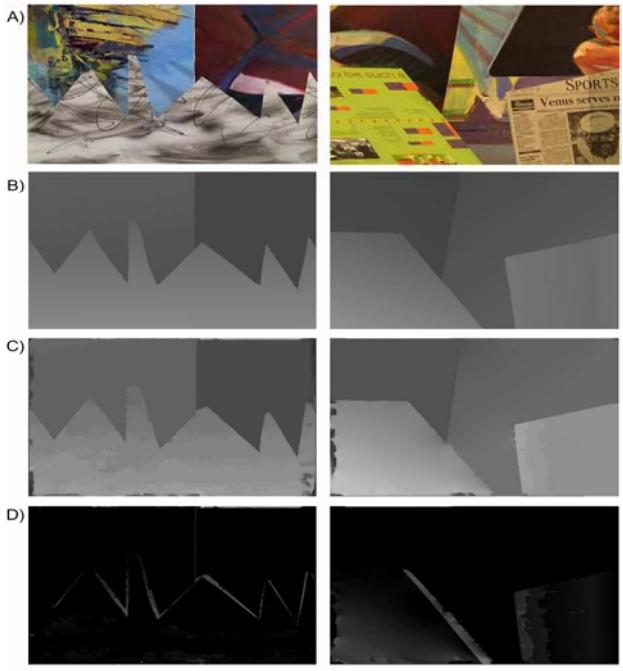

Figure 10. (A) Sawtooth image (left) and Venus image (right) stereo pair, (B) Ground truth disparity maps, (C) Estimated disparity maps, (D) Absolute disparity error.

without any postprocessing. Two popular synthetic images are chosen from the database of the University of Middlebury. The relevant disparity maps are shown in Figure 10. Furthermore an error image is also shown, in Figure 10(D), for each of the estimated disparity maps which simply is the absolute difference between the ground truth and estimated disparity maps in terms of gray scale intensity values. The absolute error can be expressed as

$$E = | d_G(x, y) - d_E(x, y) |_{\forall x \in X, \mathbb{Z}, \ y \in Y, \mathbb{Z}}, \tag{32}$$

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 which are as follows:

Table 1. Comparative performance of the proposed algorithm in terms of R, B related to the image venus

	VENUS (381×433)	
Algorithms	R	B
MW2 [12] (proposed)	1.9885	0.2262
SSD Min Filter [16]	3.7333	0.2960
Double-bp [21]	2.2114	0.2860
Symmetric-Occlusion [19]	15.7478	1.0000
Layered [22]	3.1955	0.3186

$$R = \sqrt{\frac{1}{N} \sum_{x, y} |d_E(x, y) - d_G(x, y)^2|} \tag{33}$$

and

$$B = \frac{1}{N} \sum_{(x, y)} |d_E(x, y) - d_G(x, y)^2| > \xi, \tag{34}$$

where R and B represent the *Root Mean Squared Error* (*RMSE*) and *Percentage of Bad Disparities* (*PBD*), respectively. Where N is the total number of pixels in the input image and ξ represents the acceptable deviation of the estimated disparity value from the ground truth and is taken as 1. In other words the difference between the estimated and ground truth disparity maps will be considered as bad disparity if bigger than ξ and will be accumulated to a final PBD score.

These statistics are presented in Table 1 for images *Venus* and *Sawtooth*. The multiwavelet basis used to collect the statistics is MW, taken from [12]. A number of different wavelet and multiwavelet bases are used, with different approximation order and support [3], to find the effect of these bases on the algorithm however presenting it is out of the scope of this work and will be provided in future works. For comparative performance check, RMSE and PBD of fourother algorithms, SSD min. Filter [16], Double-bp [21], Symmetric-Occlusion [19] and Layered [22], applied to the Venus image is also presented.

The maximum disparity value searched is 4 and is at the coarsest level. After that, at finer levels, the search will be single neighborhood search as can be visualized from Figure 7. The range of maximum disparity search of 4 represents 32 at the finest (input) level as can be understood by the expression 28. This supports our claim, about the proposed algorithm, for not being computationally very expensive comparative to the other local search algorithms.

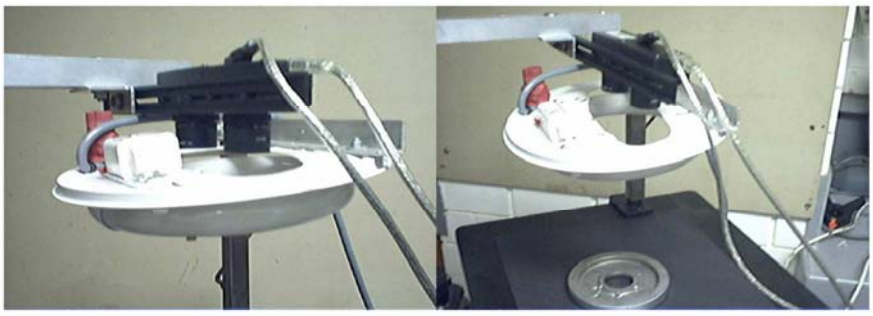

Figure 11. Reference templates extracted from the real parts.

5. Depth Estimation

The main objective of the presented work is to estimate accurate depth maps 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 [14], capable of producing accurate disparity maps by using light or laser projectors but they are both difficult and expensive to use in a typical production environment. In the presented work, only a basic hardware setup is used consisting of a pair of stereo cameras and defuse able 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.

A. Hardware setup

The complete hardware setup consists of a pair of cameras, mounted on top of a tripod stand, and a fluorescent circular light rod as shown in Figure 14. The circular light helps in keeping the light smoothly spread over the area of interest. The stereo cameras used in the work presented here are Firewire (IEEE 1394) monochrome cameras from Videre Design [6].

B. Algorithmic customization

As the algorithm is intended to be deployed in the factory environment, therefore there is a great deal of the system to be consistent, stable and fast. In order to achieve those characteristics some extra steps were brought into consideration that helps the algorithm in overcoming some of its deficiencies and in estimating accurate depth maps with higher speed, consistency and stability.

First assistance to the algorithm is the availability of the shapes of the objects that are known. For this purpose, an ideal edge map of each

object as a reference model is introduced to the algorithm. The algorithm then extracts the templates out of the reference image, which contain different shape patterns and the location of the templates with respect to each other. To keep in mind, there is only one reference model for all the objects of the same shape. An example of the reference templates for each of the two parts is shown in Figure 12.

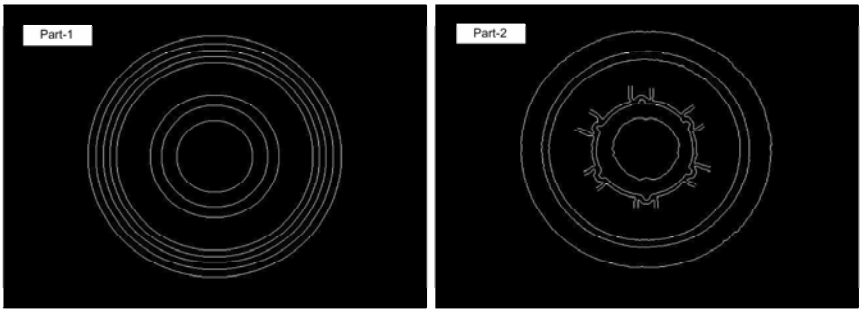

Figure 12. Reference templates extracted from the real parts.

To improve the speed of the system, algorithm tries to extract the similar templates, from the input images, as in reference model before proceeding to any coefficient to coefficient correspondence estimation. If the extracted templates are completed in shape, i.e., all pixels within the template are connected, and are very close in shape to the reference template then only few points along the pattern are taken for depth estimation considering the rest of the points will possess the similar depth. In that case a number of points, usually [20–30], are chosen randomly to extract the disparity and consequently depth. At this stage, we are making use of the assumption that any damage will change the shape of the template significantly. Therefore, if the shape of the extracted template is very similar to the one in the reference model, the template is damage free and possesses the same depth. Estimated value of disparity and depth is then assigned to the rest of the pixels belonging to the processed pattern. In that way the depth of the whole object can be estimated far quicker than by looking at each single pixel as commonly done by most of the depth estimation algorithms.

C. Estimated depth models

For performance estimation of the proposed algorithm, two different metallic parts are used as shown in Figures 13 and 15. For the validation of the estimated depth maps, between the estimated and real, one-dimensional cross-section of the depth differences along the red lines in sub-figures 13 and 15(A) is shown in sub-figures 13 and 15(E). In sub-figures 13 and 15(E) green solid line represents the estimated depth and dashed blue line represents the actual depth dimensions of the Part-1 measured manually. Whereas the red dotted line represents the difference in millimeters between the real and estimated depths. For further clarification, only depth differences (errors) are shown in Figures 14 and 16. In both Figures 14 and 16 sub-figures (A) and (B) represent the difference in depth across the red line shown in sub-figures 13 and 15(A).

In Figures 14 and 16, the sharp peaks do not represent difference/error in depth but difference in x and y dimensions. It due to fact that parts are always at different locations in front of the cameras, while image capturing, therefore appears at different locations in the image space. Furthermore the location of the parts also influence the scaling of the object in the image space as the objects can be perfectly in front of the cameras or little bit away. Algorithm works perfectly as long as the objects are completely within the view of both cameras.

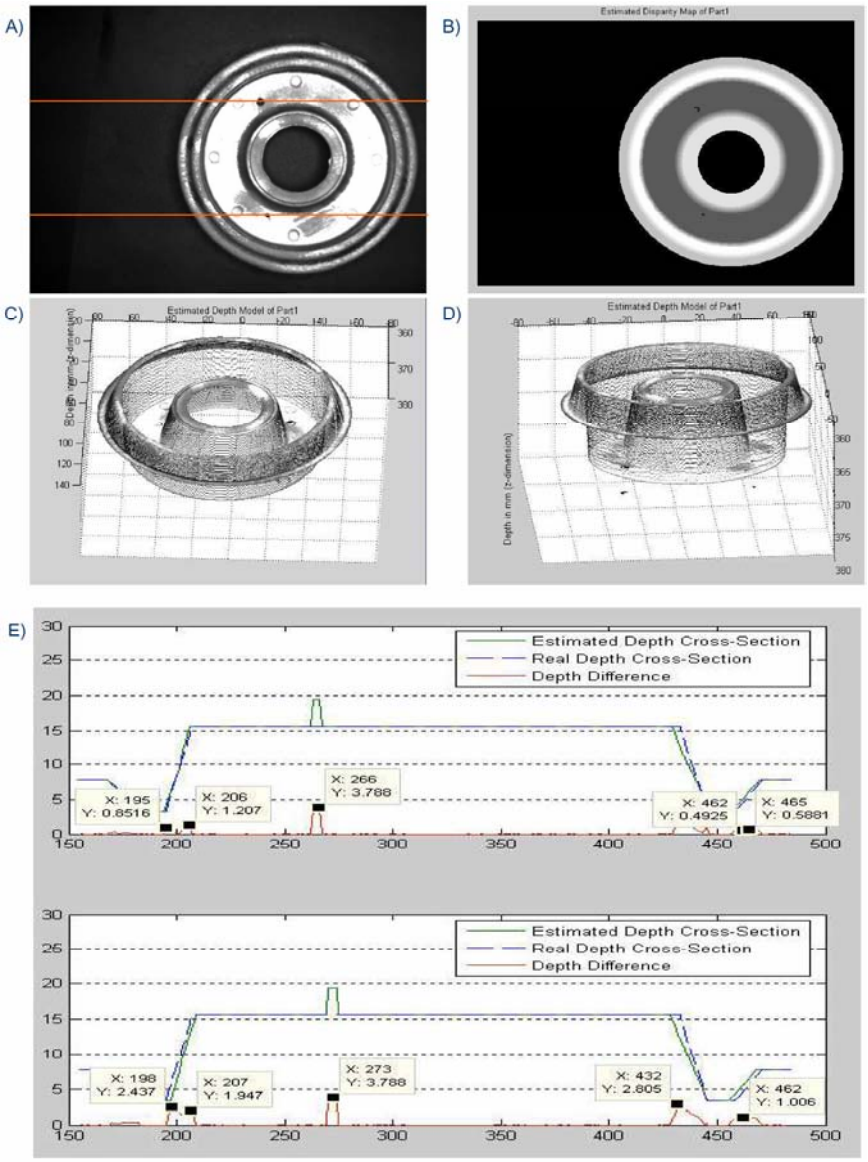

Figure 13. (A) Original image of Part-1 (Bad sample), (B) Estimated Disparity Map of Part-1, (C) Estimated 3D Depth of Part-1 in (mm), (D) Estimated 3D Depth of Part-1 in (mm) (difference view), (E) Different view of the Estimated Depth map (mm).

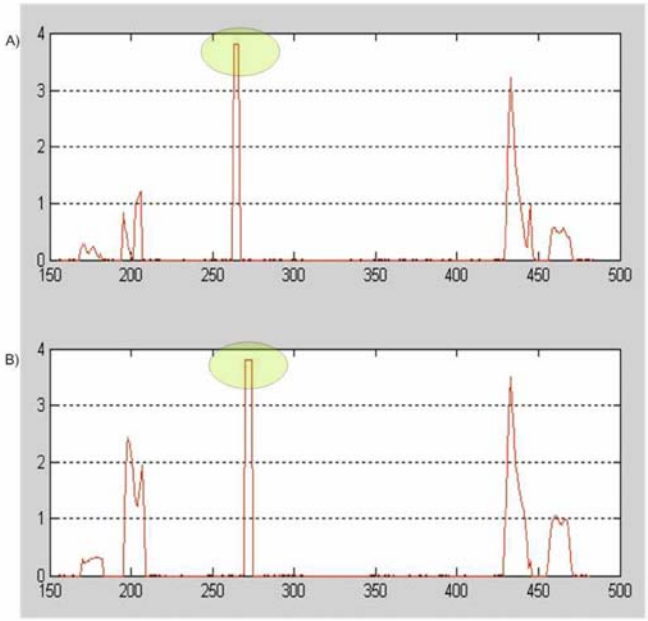

Figure 14. Part-1 (A) Difference in Depth in (mm) across the top line in Figure 13(A), (B) Difference in Depth in (mm) across the bottom line in Figure 13(A).

The statistics provided in Figures 13, 14, 15 and 16 are obtained exclusively for the sake of presentation however in real life the process is done implicitly by comparing the estimated depth with the known real depth without aligning or superimposing the estimated depth on the real one as can be seen in the above figures. Therefore, considering the above explanation, the real depth differences in Figures 14 and 16 are the peaks with flat tops as shown in Figures 14(A, B) and 16(A, B) with highlighted circles.

Referring to Figures 13 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 [0.80–1.49mm] whereas for Part-2 the difference is [0.018–1.42mm]. 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 is set for differentiating between good and defective parts in an inspection system. The time taken by the algorithm

is within the range of [10–15] seconds for the images of size [640–480] with different variations of damages.

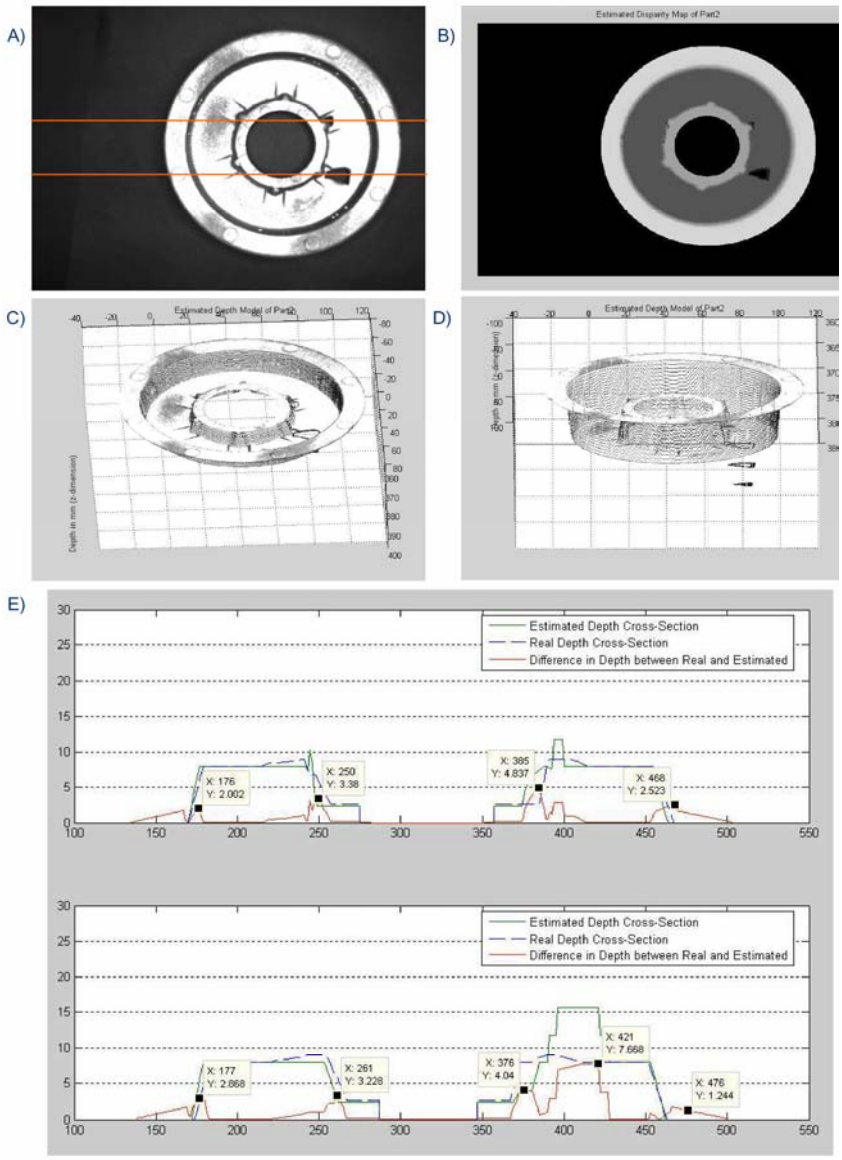

Figure 15. (A) Original image of Part-2 (Bad sample), (B) Estimated Disparity Map of Part-2, (C) Estimated 3D Depth of Part-2 in (mm), (D) Estimated 3D Depth of Part-2 in (mm) (difference view), (E) Different view of the Estimated Depth map (mm)

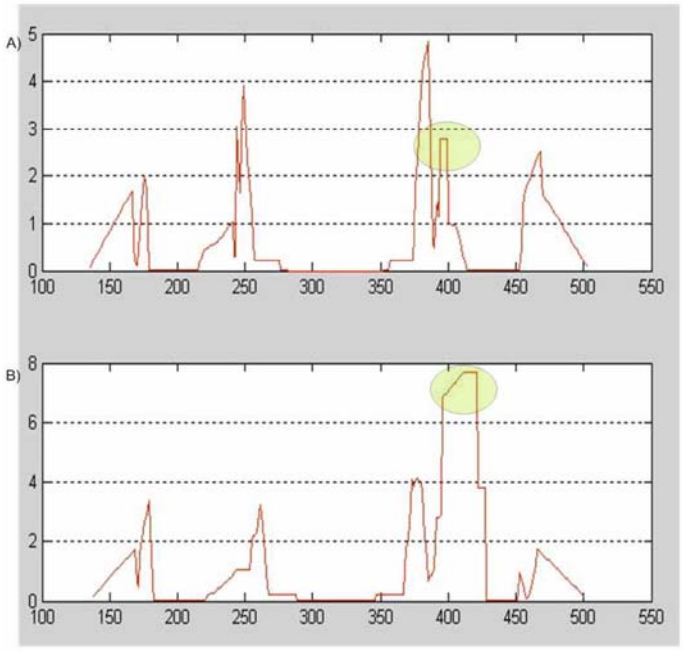

Figure 16. Part-2 (A) Difference in Depth in (mm) across the top line in Figure 15(A), (B) Difference in Depth in (mm) across the bottom line in Figure 15(A).

6. Conclusion

A novel and robust vision system is presented, 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 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 can be detected which is a major contribution towards an automated quality inspection system.

The developed vision system consists of a new proposed robust algorithm. The proposed algorithm uses the stereo vision capabilities along with multiwavelets scale space representation and hierarchical correspondence search to estimate disparity maps and the concerned 3D depths. The translation invariant multiwavelets transform modulus maxima (WTMM) are used as matching features. To keep the whole matching 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 coefficients 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 matching process is constrained to uniqueness, continuity and smoothness.

7. Acknowledgements

CAST was established and is founded in part by the Australian Government's Cooperative Research Centres Program. The authors would also like to acknowledge the support base provided by the Intelligent Systems Research Group at School of Engineering and Technology of Deakin University throughout this work.

References

[1] J. Bergen, P. Anandan, K. Hanna and R. Hingorani, Hierarchical model-based motion estimation, ECCV, (1992), 237-252.

[2] A. Bhatti and S. Nahavandi, Accurate 3D modelling for automated inspection: A stereo vision approach, Proc. of the Intelligent Production Machines and Systems - First I* Proms Virtual Conference, 2005.

[3] A. Bhatti and H. Zkaramanli, *M*-band multiwavelets from spline super functions with approximation order, International Conference on Acoustics Speech and Signal Processing, (ICASSP 2002), 2002.

[4] Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, IEEE TPAMI 23(11) (2001), 1222-1239.

[5] Changming Sun, Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques, International Journal of Computer Vision 47 (2002), 99-117.

[6] V. Design, http://www.videredesign.com.

[7] L. Di Stefano, M. Marchionni, S. Mattoccia and G. Neri, A Fast Area-based Stereo Matching Algorithm, Image and Vision Computing, 22(12) (2004), 938-1005.

[8] A. Fusiello, V. Roberto and E. Trucco, Symmetric stereo with multiple windowing, International Journal of Pattern Recognition and Artificial Intelligence, 2000.

[9] R. Hartley and A. Zisserman, Multiple View Geometry, 2nd ed., Cambridge, UK: Cambridge University Press, 2003.

[10] T. S. Huang and A. N. Netravali, Motion and structure from feature correspondences: A Review, Proc. of the IEEE. 1994.

[11] S. Mallat, A Wavelet Tour of Signal Processing, 2nd ed., Academic Press, 1999.

[12] H. Özkaramanli, A. Bhatti and B. Bilgehan, Multi wavelets from *b*-spline super functions with approximation order, Signal Processing, Elsevier Science (2002), 1029-1046.

[13] M. Pollefeys, 3D modelling from images, in Conjunction with ECCV2000: Dublin, Ireland, 2000.

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

[15] D. Scharstein and R. Szeliski, Stereo matching with nonlinear diffusion, Int. J. of Computer Vision 28(2) (1998), 155-174.

[16] D. Scharstein and R. Szeliski, A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms, Int. J. of Computer Vision 47 (2002), 7-42.

[17] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. Journal of Computer Vision (IJCV), 47 (2002), 7-42.

[18] F. Shi, N. Hughes and G. Robert, SSD matching using shift-invariant wavelet transform, British Machine Vision Conference, 2001.

[19] J. Sun, Y. Li, S. Kang and H. Shum, Symmetric stereo matching for occlusion handling, CVPR, 2 (2005), 339-406.

[20] A. Witkin, D. Terzopoulos and M. Kass, Signal matching through scale space, Int. J. of Computer Vision 1 (1987), 133-144.

[21] Q. Yang, L. Wang, R. Yang, H. Stewenius and D. Nister, Stereo matching with color-weighted correlation, hierarchical belief propagation and occlusion handling, CVPR06, 2 (2006), 2347-2354.

[22] L. Zitnick, S. Kang, M. Uyttendaele, S. Winder and R. Szeliski, High-quality video view interpolation using a layered representation ACM Transactions on Graphics (TOG), 23(3) (2004), 600-608.

Kindly return the proof after correction to: *The Publication Manager* *Pushpa Publishing House* *Vijaya Niwas* 198, *Mumfordganj* *Allahabad*-211002 (*India*) along with the print charges* by the fastest mail ***Invoice attached**	Proof read by: Dr. Asim Bhatti Signature: ..... Date: 18-08-08 Tel: +61 3 52272548 Fax: +61 3 52271046 e-mail: asimbh@ieee.org Number of additional reprints required ....NONE Cost of a set of 25 copies of additional reprints @ Euro 12.00 per page. (25 copies of reprints are provided to the corresponding author ex-gratis)