Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition, Beijing, China, 2-4 Nov. 2007

Depth Estimation using Multi-Wavelet Analysis based Stereo Vision Approach

A. Bhatti, S. Nahavandi

Intelligent Systems Research Lab, Deakin University, Vic 3217, Australia
E-MIAL: asimbh@ieee.org, nahavand@deakin.edu.au

Abstract

The problem of dimensional defects in aluminum die-casting 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 of aluminum die-castings, it is extremely hard to estimate true dimensionality of the die-casting, autonomously. 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 stereo cameras pair and a simple fluorescent light. The developed system is capable of estimating surface depths within the tolerance of 1.5 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.

Keywords: 3D depth estimation; multiwavelets transform modulus maxima;disparity estimation; stereo vision

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 well adapted to the unpredictable environments. The visual inspection processes, on the other hand, require observing the same type of image 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 [1] capable of estimating depth to reasonable 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 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 dose not require any predefined orientation and location during image capture, therefore the system can be deployed on a production line for automated inspection.

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 uses the well-known technique of coarse-to-fine matching to address the problem of stereo correspondence using the multiwavelets transform modulus maxima (MWTMM) as corresponding features. The algorithm introduced a new comprehensive selection criterion called strength of the candidate (SC) unlike many existing algorithms where selection is solely based on different aggregation costs [7, 8] 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 multiwavelet based multi-resolution analysis. Symbolic tagging procedure helps to keep the track of different candidates to

1-4244-1066-5/07/$25.00 ©2007 IEEE

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].

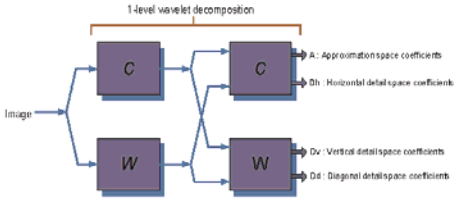

Figure 1. Mallat's dyadic wavelet filter bank

1.1 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

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

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 sub-spaces of A_{s-1}.

The use of Mallat's dyadic filter-bank [3] results in three different detail space components, which are the horizontal, vertical and diagonal. Figure 1 can best visualize the graphical representation of the used filter-bank, where L and H represents the low-pass and high-pass filters consisting of the scaling functions and wavelets coefficients, respectively.

1.2 Wavelet Transform Modulus

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

$$WTM_{s,k} = W_{s,k} \angle \Theta_{W_{s,k}} \tag{2}$$

Where $W_{s,k}$ is

$$W_{s,k} = \sqrt{|D_{h,s,k}|^2 + |D_{v,s,k}|^2} \tag{3}$$

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

$$\Theta_{W_{s,k}} = \begin{cases} \alpha(k) & \text{if } D_{h,s,k} > 0 \\ \pi - \alpha(k) & \text{if } D_{h,s,k} < 0 \end{cases} \tag{4}$$

where

$$\alpha(k) = tan^{-1}\left(\frac{D_{v,s,k}}{D_{h,s,k}}\right) \tag{5}$$

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

$$\vec{n}(k) = [cos(\Theta_{W_{s,k}}), sin(\Theta_{W_{s,k}})] \tag{6}$$

An edge point is the point p at some scale s such that $WTM_{s,k}$ is locally maximum at $k = p$ and $k = p + \varepsilon\vec{n}(k)$ 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 [3].

2 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 2, presents a detailed visual representation of the correspondence estimation algorithm.

2.1 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

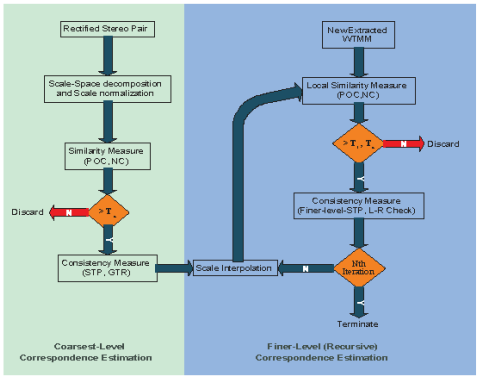

Figure 2. Block diagram of the correspondence estimation algorithm

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, which 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 that work. The objects that we are concerned with, for the depth estimation, 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 transform decomposition. It is done by dividing the details space with the approximation space and can be defined as

$$NW_{s,k} = \frac{D_{dc,s}}{|A_s|} \qquad \forall\, dc \in \{h, v, d\} \tag{7}$$

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

2.1.1 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 multi-window approach [4] is used to enhance the performance of correlation based similarity measure. The details about the improvements from single window to multi-window approach can be found in [4]. The multi-window correlation score can be defined as

$$NC(s,k) = NC_{s,k,W_0} + \sum_{j=1}^{n_W/2} NC_{s,k,W_j} \tag{8}$$

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 jth surrounding windows with n_W number of surrounding windows. In 8, 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 $[0-1]$.

2.1.2 Probabilistic Weighting

To make the correlation based selection criteria more comprehensive, a probabilistic weighting for the correlation measure in (8), is introduced. It is the probability of selection of a point, let say ith point, from within each search space, as a candidate C_i and out of r^2 search spaces as

$$P_c(C_i) = n_{C_i}/r^2 \qquad \text{where} \qquad 1 \le n_{C_i} \le r^2 \tag{9}$$

where n_{C_i} is the number of times a candidate C_i is selected and r is the multiplicity of multi-filter 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 (9) that the $P_c(C_i)$ lies between the range of $[1/r^2\ \ 1]$. 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 jth candidate C_j is selected $r^2/2$ times out of r^2 search spaces then **POC** of jth candidate is 0.5, i.e. $P_c(C_j) = 1/2$. The correlation score in expression 9 is then weighted with **POC** as

$$CS_{C_i} = P_c(C_i) \sum_{n_{C_i}} NC_{C_i}(x,d) \;\; \forall_{\, n_{C_i} \in \mathbf{Z}\, :\, n_{C_i} \le r^2} \tag{10}$$

The probabilistic weighted correlation score CS_{c_i}, in (10), 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.

2.1.3 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

$$\begin{aligned} NC_{s,k} \geq T_{c_1}, \quad \text{and} \quad P_c(C_i) = 1, &\Longrightarrow \boldsymbol{Op} \\ NC_{s,k} \geq T_{c_1}, \quad \text{and} \quad 0.5 \leq P_c(C_i) < 1, &\Longrightarrow \boldsymbol{Cd} \\ NC_{s,k} \geq T_{c_2}, \quad \text{and} \quad 2/r^2 \leq P_c(C_i) < 0.5, &\Longrightarrow \boldsymbol{Cr} \end{aligned} \tag{11}$$

It can be seen from the first expression in (11), 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 [5].

2.1.4 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 (11) 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, Matric and Euclidean [6]. 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

$$d_{Op_{n,i}} = ||\mathbf{Op}_{\mathbf{1}_i} - \mathbf{Op}_{\mathbf{2}_i}||_{n_r \,:\, n_r \leq n_{\mathbf{Op}}} \tag{12}$$

where $||.||$ defines the Euclidean distance between the pair of points with tags **Op** referring to image 1 and 2, respectively. The process in (12) 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

$$d_{\mathbf{Cd}_j} = ||C_{\mathbf{Cd},1} - C_{\mathbf{Cd},2_j}||_{m \,:\, j=1\cdots m} \tag{13}$$

where m is the number of candidates $C_{\mathbf{Cd},2_i}$ selected from second image with potential to make a pair with $C_{\mathbf{Cd},1}$ in the first image. From (12) and (13) we can define the *ADD* as

$$d_{A_{C_i}} = \left| \frac{\overline{d_{\mathbf{Cd}_i} - d_{Op_{n,\,i}}}}{d_{\mathbf{Cd}_i} + d_{Op_{n,\,i}}} \right|_n \tag{14}$$

where $d_{A_{C_i}}$ is the *ADD* for ith 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*. It is worth mentioning that absolute distances are invariant through Euclidean Transformation [5].

Before proceeding to the definition of *RDD* it is worthwhile to visualize the geometric refinement procedure as in Figure 3. 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 color, are spread all over the image and act 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**. Similarly, *RDD* can be defined by the following expression

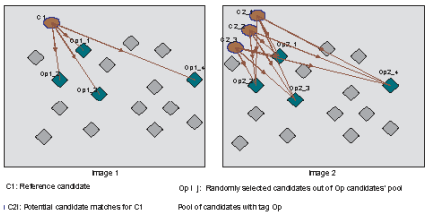

Figure 3. Geometric refinement procedure

$$d_{R_{C_i}} = \min_j \left(\left| \frac{\overline{d_{R_{C_{1,i}}} - d_{R_{C_{2,i,j}}}}}{d_{R_{C_{1,i}}} + d_{R_{C_{2,i,j}}}} \right|_n \right) \tag{15}$$

where

$$d_{R_{C_{1,i}}} = ||C_1 - \mathbf{Op}_{1,i}||_{i \,\in\, n_{\mathbf{Op}}} \text{ where } i = 1 \cdots n_r \tag{16}$$

and

$$d_{R_{C_{2,i,j}}} = ||C_{2,j} - \mathbf{Op}_{2,i}||_{i \in\, n_{\mathbf{Op}},\, j \in\, m} \text{ where } i = 1 \cdots n_r, \; j = 1 \cdots m \tag{17}$$

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

$$d_{SC_i} = \min_j \left(\left| \frac{S_{C_{1,i}} - S_{C_{2,i,j}}}{S_{C_{1,i}} + S_{C_{2,i,j}}} \right|_n \right) \quad (18)$$

The term $(.)_n$ defines the average over n repetitions, where n is usually taken within the range of [3 5]. Using (14), (15) and (18) 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 (10) to make the geometric measure more comprehensive as

$$Gc_i = max_i(CS_{C_i} (\overline{e^{-d_{AC_i}} + e^{-d_{RC_i}} + e^{-d_{SC_i}}})) \quad (19)$$

The expression in (19) 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**.

2.1.5 Scale Interpolation

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

$$d_F = d_L + 2d_c \quad (20)$$

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 4.

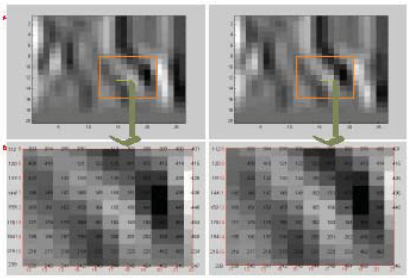

Figure 4. An outcome of coarsest level matching

2.2 Finer-Level Correspondence Estimation

The coarsest-level matching process, from section-A and Figure 2, results in a number of matched pairs, which need to be interpolated to the finer level, i.e. to the $N-1$th level. The search for the best match, performed at the finer level, is a local search that is only defined at the predefined locations where the matches are found earlier at the coarser level. For this search process Similarity Measure Block, shown in Figure 2 and presented in sections **A-1** to **A-2**, is used. After that step, all candidates are assigned with CS as defined in (10). To address the problem of ambiguity, left-right consistency check on discrete disparity is performed which can be expressed as

$$-d_{i,1}(x,y) = d_{i,2}(x + d_{i,1}(x,y), y) \quad (21)$$

Where $d_{i,j}(x,y)$ is the estimated discrete disparity related to ith candidate in the jth image.

3 Depth Estimation

For performance estimation of the proposed algorithm, two different aluminum die-castings 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 Figure 6. One-dimensional cross-section of the depth difference is 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.

Referring to Figure 5 and Figure 6, 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 can be set for differentiating between good and defective parts in an inspection system. Moreover, in Figure 5-E and Figure 6-E the sharp peaks are in fact due to the difference in x and y dimensions rather than the difference in depth.

4 Conclusion

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

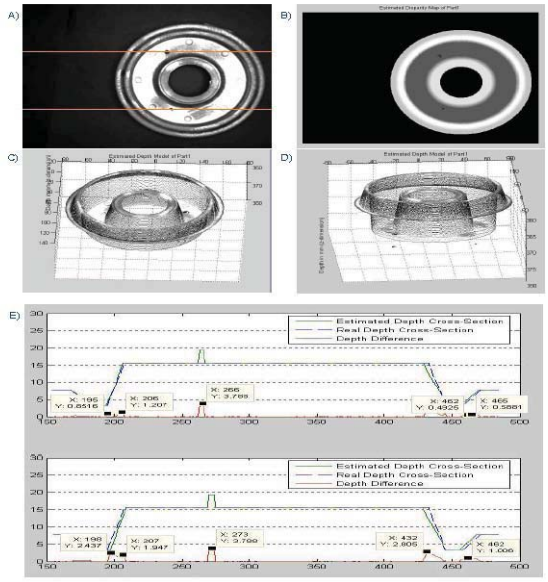

Figure 5. A: Original image of Part-1 (Bad sample) B: Estimated Disparity Map of Part-1 C: Estimated 3D Depth of Part-1 in (mm) D: Estimated 3D Depth of Part-1 in (mm) (difference view) E: Different view of the Estimated Depth map (mm)

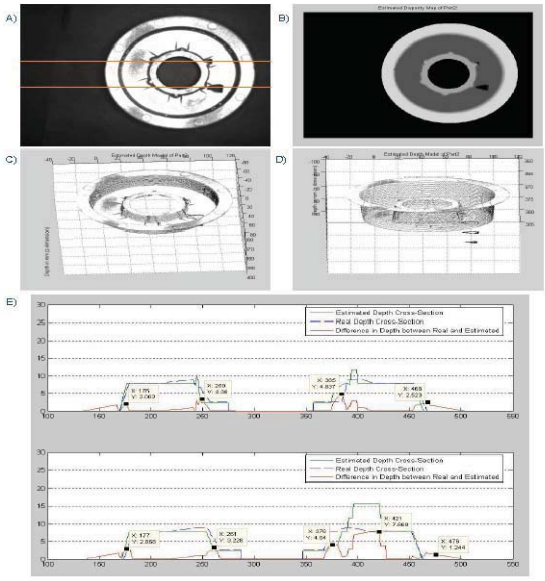

Figure 6. A: Original image of Part-2 (Bad sample) B: Estimated Disparity Map of Part-2 C: Estimated 3D Depth of Part-2 in (mm) D: Estimated 3D Depth of Part-1 in (mm) (difference view) E: Different view of the Estimated Depth map (mm)

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.

References

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

[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] S. Mallat, *A Wavelet Tour of Signal Processing*, Vol. 2nd edition: Academic Press, 1999

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

[5] R. Hartley and A. Zisserman, *Multiple View Geometry*, Vol. Second Edition, Cambridge, UK: Cambridge University Press, 2003

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

[7] F. Shi, N. Hughes, and G. Robert, *SSD Matching Using Shift-Invariant Wavelet Transform*, British Machine Vision Conference, 2001.

[8] Changming Sun, *Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques* International Journal of Computer Vision, 47: p. 99-117, 2002