AREA BASED STEREO IMAGE MATCHING TECHNIQUE USING HAUSDORFF DISTANCE AND TEXTURE ANALYSIS

Jyoti Joglekar [a, *], Shirish S. Gedam [b]

[a] CSRE, IIT Bombay, Doctoral Student, Mumbai, India – jyotij@iitb.ac.in
[b] Centre of Studies in Resources Engineering, IIT Bombay, Associate Professor, Mumbai, India – shirish@iitb.ac.in

Working Group III/4

KEY WORDS: Area, Image, Matching, Texture, Hausdorff distance, Reconstruction

ABSTRACT:

A conventional image matching techniques may be classified as either area based or feature based methods. In this paper an area based image matching method is proposed for dense disparity map. The method is a composite technique where first the similarity measure between template window and search window is found by normalized cross correlation technique. Few best matches are selected for the template window from the search sub windows, considering the largest normalized cross correlation coefficient. Further edge map is obtained for stereo image pair using canny edge detector. The matches for the template window are filtered using Hausdorff distance technique. Further texture analysis of the same template window and selected search windows is the third measure to decide the accurate match. Texture analysis is done with the co-occurrence matrices which is a two dimensional histogram of the occurrence of pair of intensity value in a given spatial relationship. With this composite method dense point to point correspondence can be achieved with greater accuracy. This method is tolerant to radiometric distortions and parallel processing of the three techniques will improve the speed.

1. INTRODUCTION

Stereo Image matching is one of the core research areas in Computer Vision and digital Photogrammetry. Technological developments in stereo image matching have advanced from the primitive area based cross correlation technique to more and more precise feature based matching. Stereo allows us to recover information from the given two images about a three dimensional location of objects, which does not exist in any single image. The main goal of stereo image matching is to recover depth information from the given two or multiple images (Zitova and Flusser, 2003).

In order to recover depth information the stereo images should be brought into point-point correspondence. Correspondence points are the projections of a single point in to the three-dimensional scene. The difference between the locations of these two correspondence points is known as parallax or disparity which is a function of position of the point in the scene, orientation and physical characteristics of the camera. So epipolar line and disparity can be used as constraints for matching. Though feature based technique is more accurate, it gives sparse disparity map. We are proposing an area based method as it gives dense disparity map for 3D reconstruction.

The techniques used are zero mean normalized cross correlation, Hausdorff distance and texture analysis. In this method the basic units that are used for matching is regularly sized neighbourhood of a pixel. The position of the given pattern is determined by a pixel wise comparison of the image with a given template that contains the desired pattern. For this the template is shifted m discrete steps in x direction and n discrete steps in the y direction of the image and the comparison is calculated over the template area for each position (m, n).

Dense depth measurements are required in applications such as teleconferencing, robot navigation and control. As the taxonomy given in (Peleg, and Weiser, 1996), stereo algorithms that generate dense depth measurements can be roughly divided in to two classes, namely global and local algorithms. Global algorithms (Fusiello, Roberto, and Trucco, 2000) based on iterative schemes that carry out disparity assignments on the basis of the minimization of a global cost function.

As compared to local algorithm, global algorithm yield accurate and dense disparity measurement but computational cost of these algorithms is not suitable for real time application.

Local algorithms (Muhlmann, Maier, Hesser and Manner, 2002; Fusiello, Trucco and Verri, 2000; Trucco and Verri,1998) also referred as area-based algorithms calculate the disparity at each pixel based on photometric properties of the neighboring pixels compared to global algorithm, local algorithms yield significantly less accurate disparity maps. So to improve the accuracy the composite algorithm is proposed here. By selecting the disparity range found out by visual inspection of stereo image pair the search area can be limited and the computational speed can be improved so that the matching algorithm can be used for real time applications.

Area based method uses pixel intensity directly to compute the similarity measure between a small template and a large search window by the quantities of statistical correlation, Fourier cross power spectrum or mutual information. The area based methods merge the feature detection step with the matching part. These methods deal with the images without attempting to detect salient objects. There is a high probability that a window containing a smooth area without any prominent details will be matched incorrectly with other smooth areas in the reference image due to its saliency.

Classical area based methods like cross correlation exploit for matching directly image intensities without any structural analysis. Consequently they are sensitive to the intensity changes, introduced by noise, varying illumination and/or by using different sensor types.

There are generalized versions of cc for geometrically more deformed images. They compute the cc for each assumed geometric transformation of the sensed image window. Similar to the cc methods is the sequential similarity detection algorithm SSPA (Egnal and Wildes, 2000). It uses the sequential search approach and a computationally simpler distance measure than the cc. it accumulates the sum of absolute differences of the image intensity values. The method is likely to be less accurate than the cc but it is faster.

Two main drawbacks of the correlation like methods are the flatness similarity measure maxima and high computational complexity. In the proposed algorithm reducing the search window by finding disparity range with visual inspection the search area is limited. Hence the computational complexity is reduced. Further structural analysis of the probable matches is done with Hausdorff distance criteria and texture analysis to improve accuracy of the match.

The Section 2 explains three techniques used in the proposed algorithm. Section 3 explains the proposed algorithm, which is a composite technique using zero mean normal cross correlation, Hausdorff distance and texture analysis. In section 4, results are presented for a sample stereo image pair. Finally the analysis of the result and an outlook to future research activities is presented.

2. TECHNIQUES USED FOR IMAGE MATCHING

2.1 Zero mean normalized cross correlation algorithm

The problem treated in this paper is to determine the position of a given pattern in a two dimensional image f. Let f(x, y) denote the intensity value of the image f of size M x N at the point (x, y), x $\in$ { 0, 1, ……, M-1}, y $\in$ {0,1, ……,N-1}. The pattern is represented by a given template t of size (p x q). A common way to calculate the position (m_{pos}, n_{pos}) of the pattern in the image f is to evaluate the zero mean normal cross correlation value ρ, at each position (m, n) for f and template t which has been shifted m steps in x direction and n steps in y direction. The size of search window which is larger than template window is determined by visual inspection of the stereo image pair with horizontal and vertical disparity.

Equation (1) gives a basic definition for the zero mean normal cross correlation coefficient.

$$\rho = \frac{\sum\limits_{x,y}(f(x,y) - \bar{f}_{m,n})(t(x-m, y-n) - \bar{t})}{\sqrt{\sum\limits_{x,y}(f(x,y) - \bar{f}_{m,n})^2 \sum\limits_{x,y}(t(x-m, y-n) - \bar{t})^2}} \qquad (1)$$

In Equation (1) $\bar{f}_{m,n}$ denotes the mean value of $f(x, y)$ within the area of the template t shifted to (m, n) which is calculated as

$$\bar{f}_{m,n} = \frac{1}{p \times q} \sum_{x=m}^{m+p-1} \sum_{y=n}^{n+q-1} f(x,y) \qquad (2)$$

With similar notation $\bar{t}$ is the mean value of the template t. The denominator in equation (1) is the variance of the zero mean image function $f(x,y) - \bar{f}_{m,n}$ and the zero mean template function $t(x-m, y-n) - \bar{t}$. Due to this, the zero mean normalized cross correlation coefficient at (m,n) is independent to changes in brightness or contrast of the image which is related to mean value and standard deviation.

The desired position (m_{pos}, n_{pos}) of the pattern which is represented by t is equivalent to the position (m_{max}, n_{max}) at maximum value ρ_{max} of ρ(m, n). The zero mean normalized cross correlation method is more robust than other similarity measures like simple covariance or sum of absolute difference.

2.2 Hausdorff distance based image comparison

Hausdorff distance measures the extent to which each of model set lies near some point of an image set. The distance can be used to determine the degree of resemblance between two objects that are superimposed on one another. In this paper the second measure to improve the score of a match of template window to the search sub window.

A central problem in pattern recognition and computer vision is determining the extent to which one shape differs from another. Template matching can be viewed as a technique for determining the distance between shapes. In this algorithm Hausdorff distance is used as one of the measure to decide the relative position of a model and an image. The Hausdorff distance is a max-min distance as defined below.
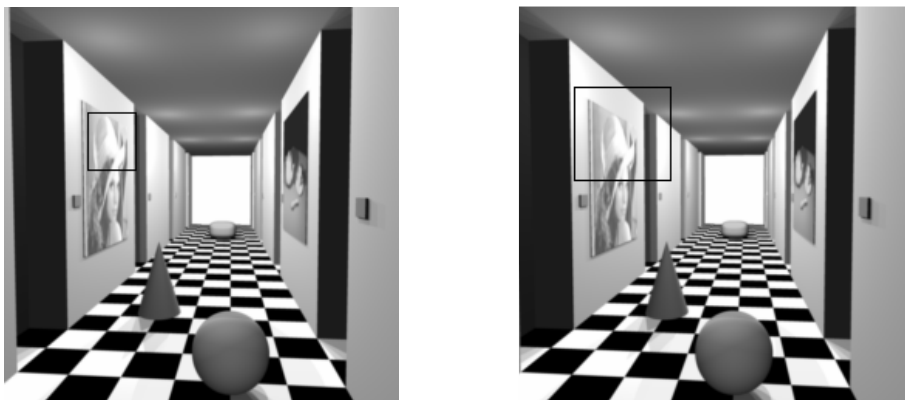

Figure 1 : Stereo image pair with left image with template window and right image with search window

Given two finite point sets A={ a_1, a_2, ……,a_n} and B= { b_1, b_2, ……,b_n}, the Hausdorff distance is defined as

$$H(A,B) = \max(h(A,B), h(B,A)) \qquad (3)$$

Where

$$h(A,B) = \max_{a \in A} \min_{b \in B} \| a - b \| \qquad (4)$$

And $\| . \|$ is L_2 or Euclidian norm. the function h(A,B) is called the direct Hausdorff distance from A to B. it identifies the point $a \in A$, that is farthest from any point of B and measures the distance from a to its nearest neighbour in B. Distance of each point of A to the nearest point of B is used for deciding ranking of each point of A. Thus the largest ranked point is considered as the distance (the most mismatched point of A). That means if h(A,B) = d then each point of A must be within distance d of some point of B and there is some point of A that is exactly distance d from nearest point of B (the most mismatched point). Hausdorff distance H(A,B) is the maximum of h(A,B) and h(B,A). It measures the degree of mismatch between two sets by measuring the distance of the point A that is farthest from any point of B. In this method of comparing shapes there is no explicit pairing of points of A with points of B. The computation time for H(A,B) will be o(mn) for two sets of size m and n.

2.3 Texture analysis using co-occurrence matrices

The co-occurrence has been described in image processing literature as gray-tone spatial dependence matrices (Haralick, Shanmugam and Dinstein, 1973). It has been widely used in texture analysis. The co-occurrence matrix in it core is a two-dimensional histogram of the occurrence of pairs of intensity values in a given spatial relationship.

Co-occurrence matrices are based on the relative frequencies $p(i, j)$ with which two pixels with a specified separation occur in the image, one with gray level i and another with gray level j. The separation is usually specified by distance vector s (d, θ). Pixel distance d and angular orientation θ are parameters of a particular co-occurrence matrices. Using different parameters, several matrices can be derived as shown in figure 2. The matrices obtained are then used for extraction of texture features. For the image of N gray levels, co-occurrence matrices cm can be obtained by estimating the pair wise statistics of pixel intensity. The size of cm is determined by the number of gray levels N in the input image.

The matrices cm are a functions of the angular relationship between the pixels as well as a function of the distance between them. An example of a 4x4 image with four gray levels and the computation of the co-occurrence matrices for d=1 and θ varying from 0° to 135° by 45° is given in figure 2.
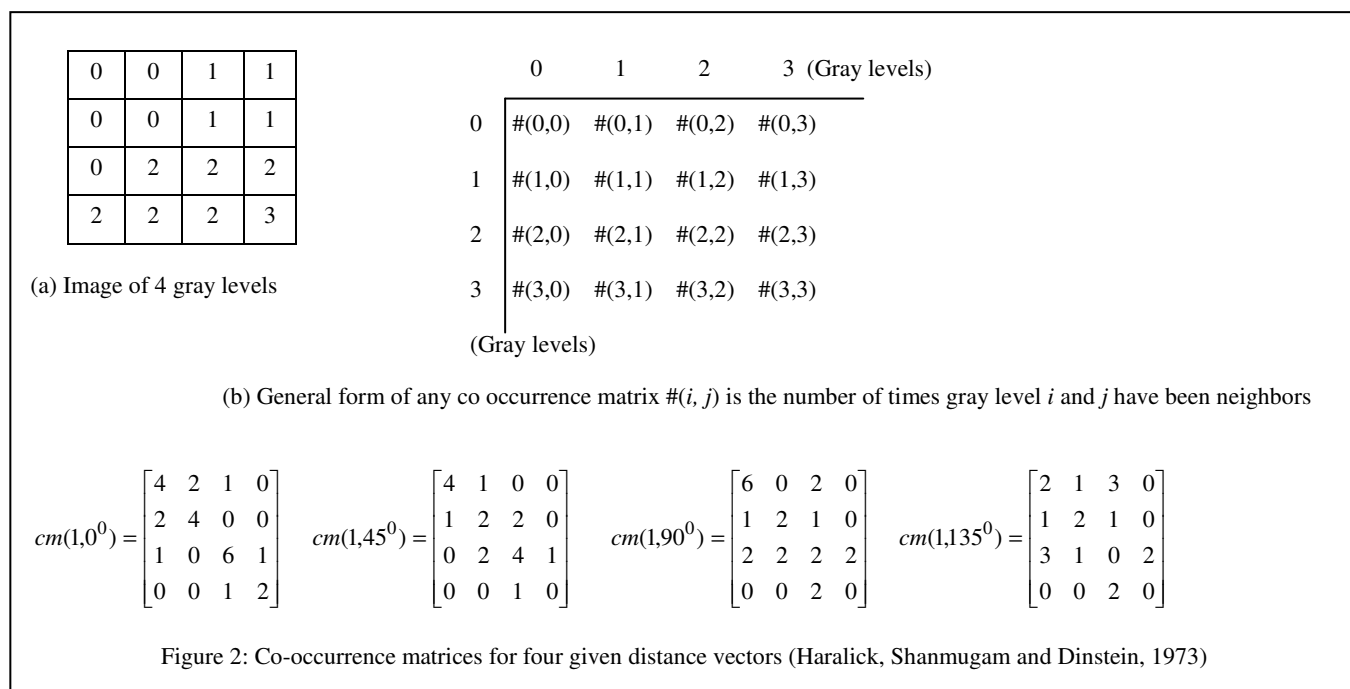
In texture classification, individual elements of the co-occurrence are rarely used. Instead, features are derived from the matrix. A large number of textural features have been proposed starting with the original fourteen features described by Haralick, however only some of these are in wide use. The features which are used are listed as following:

Inverse Difference Moment

$$f_1 = \sum_{i,j} P(i, j)/1+ (i- j)^2 \qquad (5)$$

Contrast

$$f_2 = \sum_{i,j} \delta_{ij}^2 P(i, j) \qquad (6)$$

0	0	1	1
0	0	1	1
0	2	2	2
2	2	2	3

(a) Image of 4 gray levels

	0	1	2	3 (Gray levels)
0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(Gray levels)

(b) General form of any co occurrence matrix #(*i, j*) is the number of times gray level *i* and *j* have been neighbors

$$cm(1,0^0) = \begin{bmatrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \quad cm(1,45^0) = \begin{bmatrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad cm(1,90^0) = \begin{bmatrix} 6 & 0 & 2 & 0 \\ 1 & 2 & 1 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \quad cm(1,135^0) = \begin{bmatrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$$

Figure 2: Co-occurrence matrices for four given distance vectors (Haralick, Shanmugam and Dinstein, 1973)

Entropy

$$f_3 = -\sum_{i,j} P(i,j) \log P(i,j) \qquad (7)$$

Correlation

$$f_4 = \sum_{i,j} (i - \mu_i)(j - \mu_j) P(i,j) / \sigma_i \sigma_j \qquad (8)$$

Energy (angular second moment)

$$f_5 = \sum_{i,j} P(i,j)^2 \qquad (9)$$

μ_i and μ_j are the means and σ_j and σ_j are the standard deviations of i and j respectively. These texture features are used for image matching by comparing the feature value of template window with the search window under consideration.

The five texture features are computed from the co-occurrence matrices of the template window and the search sub window and the difference between the feature coefficients is compared with a threshold which is empirically designed out for each feature. If the difference of the feature coefficients of three features are within the predefined range then the match is accepted. Sometimes cross-correlation technique fails to compare areas of smooth variance of intensities. Hence the accuracy of the match can be improved by texture analysis technique.

3. PROPOSED ALGORITHM

A composite method for area based image matching is proposed below:

1 Zero mean normal cross correlation coefficient between template window and search sub window is computed. The sub window having maximum cross correlation coefficient is considered as more perfect match for the template window under consideration. Best five matches are selected based on coefficient value and ranking of the matches is done based on cross correlation coefficient.

2 Further, edge map of template and search sub window is obtained by canny edge detector. Thus binary edge map of template and search sub window is derived.

3 The Hausdorff distance between the binary edge map of template and search sub window is used further for refining the best matches ; the lower the distance the best the match. The ranking of the best five matches of step one is done based on Hausdorff distance criteria.

4 Further texture analysis of the template and search sub window is done using co-occurrence matrix. The texture features are derived from the co-occurrence matrix using equations (5) to (9)

5 The steps 1 to 4 are performed for every basis images which are considered as separate template window.

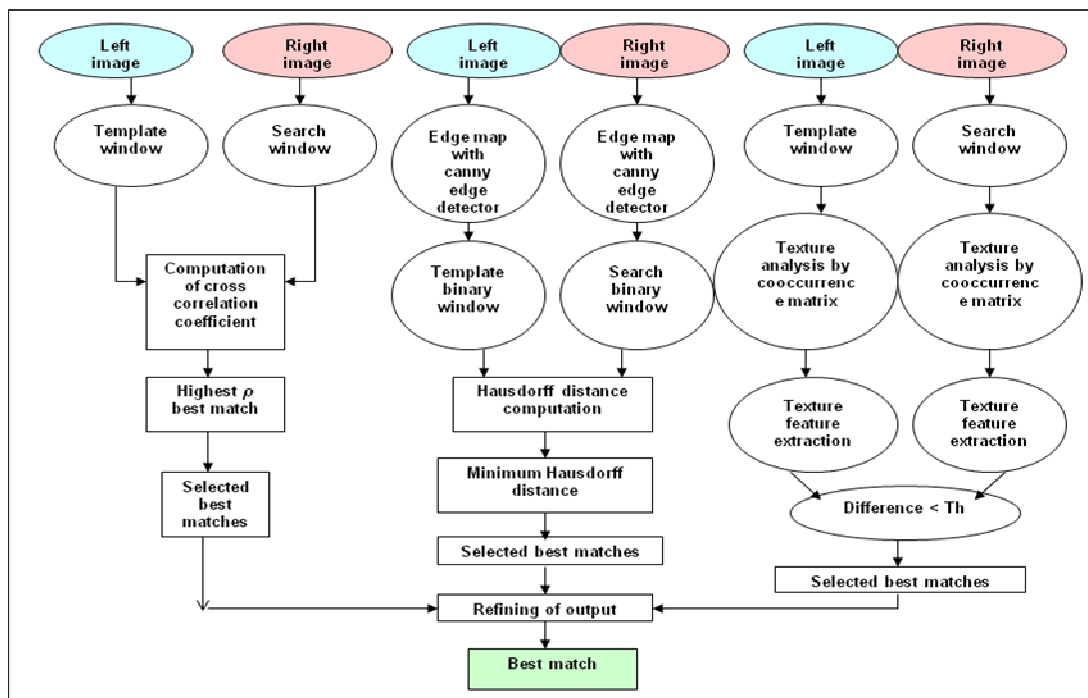Figure 3 shows the block diagram for the proposed algorithm

.

Figure 3 Block diagram for the proposed algorithm

4. RESULTS AND DISCUSSION

The proposed method is executed on the test stereo image pair shown in figure 1. As shown in the figure the left image shows the template window and the right image shows the larger search window. The size of the search window is decided by considering horizontal and vertical disparity measured by visual inspection of stereo image pair in figure 1.

As shown in figure 3, the block diagram for the proposed algorithm, the three methodologies used are zero mean normal cross correlation, Hausdorff distance and texture analysis. Parallel processing of the three methodologies can be incorporated for improving speed. Using zero mean normal cross correlation and considering largest five coefficient for the best matches for the given template window as shown in figure 4a. The matches found are shown in figure 4b. The minimum threshold of 0.8 is used for selecting the search sub window as a match. If the normalized cross correlation coefficient is less than 0.8 for all search sub window means that there is no match in the right image for the template window under consideration in the left image.

Further on the selected best matches and template window canny edge detector is applied and a set of binary images are obtained as shown in figure 4c. Computation of Hausdorff distance and ranking the matches based on Hausdorff distance is done. After the third step of the algorithm the best matches are filtered to the output as shown in figure 4d. The third refinement is done with texture analysis. Texture analysis is done on template window and selected search sub window of the figure 4d. The fourteen texture features are defined by (Haralick, R. M., Shanmugam, K. S., and Dinstein, I., 1973).

Figure 4a: Template window from left image from stereo image pair

Figure 4b: Selected sub widows as probable matches

Figure 4c: Binary edge images of sub windows of figre 4b

Figure 4d: Selected sub windows as matches for template window atfter applying Hausdorff distance criteria

Figure 4e: Best match selected after texture analysis

Out of the fourteen texture features the features entropy, contrast, inverse difference moment, correlation and energy (angular second moment) are computed and then best match is selected by comparing the feature values of template window and the refined output as shown in figure 4d. The best match is shown in figure 4e.

This procedure is repeated for every template window considered as a basis image. The first step is processed with a time complexity O(mn), where search window has pixels m X n. For further pass the comparison is done only on selected sub window making the execution faster.

The method removes most of the outliers and the accuracy of matching is improved compared to any single method giving a dense disparity map useful for 3D reconstruction. If parallel processing is done the method can be used in real time application like robotic vision. Figure 4e shows the best match.

5. CONCLUSION

Area based image matching algorithm offers a dense disparity map which is needed for 3-D construction of a scene from stereo images. In the proposed composite technique of image matching, three methodologies are used for refining the accuracy of the match. The normalized cross correlation coefficient is used for selecting best five matches and scores are given to them according to correlation coefficient value.

Further the scores of the matches are improved by Hausdorff distance technique. The third method that does texture analysis of template window and corresponding search sub window, improves the score of the matches. The best match is the one which is having highest score. This composite method of area based image matching improves the accuracy of the match and reduces number of outliers and gives dense disparity map. It can be used for real time application as the size of search window can be reduced by disparity range computed by visual inspection of stereo image pair reducing the computational cost. Further, refining of match with Hausdorff distance method and texture analysis method is done in linear time.

6. REFERENCES

Bernard, S.T. and Fischler. M.A. "Computational Stereo", *ACM Computing Surveys*, vol, 14, pp.553-572, 1982.

Canny, J., "A Computational Approach to Edge Detection",*IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp.679-714, 1986.

Egnal G., Wildes R., "Detecting binocular half-occlusions: empirical comparisons of four approaches", *Proc. Int. Conf. Comput. Vision Pattern Recognit.* (2) (2000) 466–473.

Fusiello A., Roberto V., and Trucco E., "Symmetric stereo with multiple windowing, *Int. Journal of Pattern Recognition and Artificial Intelligence,* 14:1053-1066, 2000.

Fusiello A., Trucco E., Verri E., A compact algorithm for rectification of stereo pairs, Machine Vision Appl. 12 (1) (2000) 16–22.

Haralick, R.M. and Shapiro L.G., "Computer and Robotic Vision", in *MA: Addition-Wesley,* Volume: 2, Chapter: 16, 1993

Haralick, R. M., Shanmugam, K. S., and Dinstein, I., "Textural Features for Image Classification", *IEEE International Conference on Systems, Man, and Cybernetics,* Vol. SMC-3, No.6, pp 610-621, 1973

http://www-cgrl.cs.mcgill.ca/~godfried/teaching/cg-projects/98/normand/main.html

Huttenlocher D. P., Kedam K "Efficiently computing the Hausdorff distance for point sets under translation,", *Proc. Sixth ACM Symp. Computat. Geometry,* 1990, pp.340-349.

Lucas B. D., Kanade T "An Iterative Technique with an application to Stereo Vision", *IJCAI.* 1981.

Muhlmann K., Maier D., Hesser J., Manner R., "Calculating dense disparity maps from color stereo images, an efficient implementation",, *Int. J. Comput. Vision* 47 (1–3) (2002) 79–88.

Peleg A Weiser U., MMX technology extension to the intel architecture, *IEEE Micro* 16 (4) (1996) 42–50.

Pratt, W.K , "Correlation Techniques of Image Registration.", *IEEE Transactions on Aerospace and Electronic Systems,* vol.10, no. 3, pp. 353-358, 1974.

Sun, C., "A Fast Stereo Matching Method", *Digital Image Computing: Techniques and Application,* pp. 95-100, Massey University, Auckland, New Zealand, 1997.

Trucco E.,Verri A., "Introductory techniques for 3D computer vision", *Prentice Hall, Englewood cliffs, NJ*, 1998.

Zitova, B. and Flusser J, "Image Registration Methods: A Survey", *Image and Vision Computing,* Vol 21, No. 11, pp 977-1000, 2003.