# DISPARITY MAP CALCULATION THROUGH EPIPOLAR LINES ESTIMATION FOR 3D FACIAL RECONSTRUCTION

Abbas Cheddad, Halim Setan and Zulkepli Majid

Medical Research Imaging Group (MIRG), Faculty of Geoinformation Science & Engineering
Universiti Teknologi Malaysia
e-mail: cheddad1@hotmail.com, halim@fksg.utm.my

**Abstract**

*In this paper, we tackle the problem of 3D reconstruction of human faces from a given stereo pair 2D instances, namely left and right images. The generated 3D model is developed to be the main input for Medical imaging tasks, although the model can be exploited in other fields as well. The algorithm is decomposed into two phases. The first phase deals with the location and extraction of human face from its cluttered background, this process is to focus on the region of interest (ROI) and to lessen down the computational burden. Since the input 2D pair of images are in color form, a wise decision is made to exploit their RGB color map information such that the RGB matrix is transformed into a new mapping encapsulating the intensity values (Y), the Chromatic blue ($C_b$) and Chromatic red ($C_r$) (e.g.: RGB➔$YC_bC_r$). The second phase is the core of this paper. In stereo vision knowledge of epipolar lines (ELs) is extremely important as they describe the geometric relationship between the world points and their projections on the imaging sensors. In other words, these ELs with the predetermined two epipoles will solve what is known as the correspondence problem in an efficient way and reduces the search from the original 2D plan into merely 1D vector. In fact, the ELs are prone to errors; therefore, they must be constructed with care. To eliminate this problem and to increase the precision we chose to start with an initial eight matched points to generate the Fundamental Matrix (FM), this is called the 8- Points Algorithm. The FM has a compact description of the camera parameters and it is a 3x3 matrix. The 8- Points are selected from applying Harris corner detector. Thus, the epipolar line function will be given by: EL=F*f(x,y), where f(x,y) holds the point coordinates in the form $[x,y,1]^T$, the superscript ($^T$) denotes matrix transpose. This indicates that the epipolar constraint can be established with no prior knowledge of the stereo parameters. A dense disparity map is generated to be the input for the depth calculation which, in its turn leads to 3D model formation.*

***Key Words:*** *Stereo vision, RGB Transformation, Epipolar Line, Epipolar Consstraint, Epipoles, Eight Points Algorithm, 3D Face model.*

## 1.0 INTRODUCTION

Since decades and scientists are trying to emulate human beings functionality (e.g.: sensors). This phenomenon created a fierce competition in research areas related to the above. The most recent technology is dealing with the so-called 3D **re**construction, which was initiated on the basis understanding and observation of human eyes functionality. One can conceptualize this idea by picturing human eyes as a stereo cameras and the brain is a CPU (interpreter device) as shown in Figure 1. Most of humans have the stereo imaging encapsulated in their vision system, where the two eyes are able to compute the relative distances of objects, which we refer to as *depth calculation* in this study.

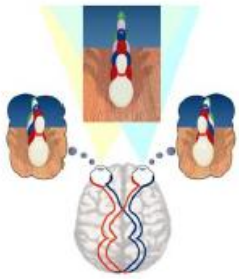


Figure 1: The concept of stereo imaging is based on the human eyes functionality

The 3D reconstruction of objects surfaces is a special discipline in computer vision, which is directed towards the recovery of 3D object shapes or the estimation of the distances between the sensors. 3D reconstruction is based mainly on the information derived from the data acquisition via sensors. One small note here; that we would like to highlight, is the prefix (re) in the word reconstruction. Definitely, we mean what we say here because the object's 3D shape is projected into two images of 2D information, while the depth information (Z coordinate) is lost during the acquisition and it is recovered through reconstruction process. This process generally goes through certain procedures such as image pre-processing (to enhance and prepare the image for further processing) and pattern analysis (to segment the image and extract its features …etc).

The applicability of this technology can be seen in different scientific fields such as computer modeling (in games technology, architecture, mechanical engineering and surgery), robotics (vehicle control), quality control (products surface inspection), locating 3D objects (automated assembly) and navigation…etc. For example, Okutomi and Katayama (2001) have shown a very interesting application to 3D reconstruction in the field of image Synthesis, Ansari and Abdel-Mottaleb (2003) and Jiang et al. (2005) exploit the 3D modeling in face recognition, while Wünstel and Schumann (2002) developed a module for automatic 3D reconstruction of the ocular fundus which was used to analyze the changes of the ocular fundus of the glaucoma patients

Stereo vision is a popular technique developed for recovering 3D structure from images. However, stable and precise estimation of disparities or depths is not a trivial task and many methods have appeared since then to address this issue.

This paper highlights a part of a huge multi-disciplinary research initiated by Universiti Teknologi Malaysia (UTM), Standards & Industrial Research Institute Malaysia (SIRIM), and Universiti Sains Malaysia (USM). The top level research focuses on the development of surgical planning system for craniofacial reconstruction, for both the soft and hard tissues (Halim et al., 2004).

## 2.0 STEREO CAMERA SET-UP

We use the cameras setup of Figure 2; which is pictured in a real indoor rig in Figure 3, to capture two near-frontal view images of the face namely left and right views. Assuming perspective projection, the 3D points are projected on the perspective rays passing through two corresponding projection centers COP1 and COP2, which are the camera centers separated by distance b from each other. The focal length f is the distance of the image planes from each center of projection.

An assumption is made here to avoid the so-called occluded scene. Hence, both views (left and right) must project same regions. In other words, the reconstruction process will be restricted to the intersection area of the two views. If for example, a set of points are viewed by one camera and not accessible by the other, this will prone these points to fail in 3D space reconstruction. Figure 4 gives a better illustration to the said problem. In the figure, it appears clearly that the two areas indicated by green arrows cannot be projected into the two cameras.

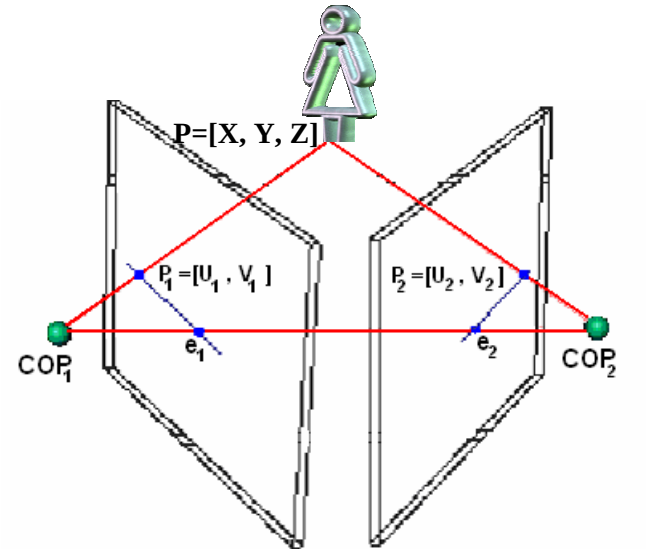



Figure 2: The topology of the stereo rig

Figure 3: The real stereo camera configuration. The white shining dots are used for camera calibration

In the above figure, a 3D point in the real world denoted as P=[X, Y, Z] is projected into a 2D points $P_1$ and $P_2$ in the left and right image respectively. The two points at which the line through the centers of projection of each image intersects the image ($e_1$ and $e_2$) are called the epipole points. The line connecting these points with any projected point in the image plan is known as epipolar line. A detailed discussion about the epipolar lines construction will be discussed later.

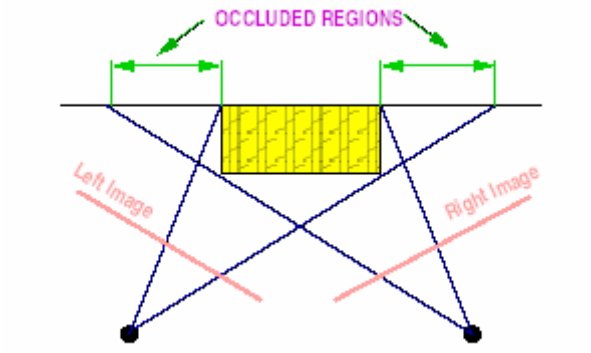



Figure 4: A sketch highlighting the problem of occluded regions

## 3.0 METHODOLOGY

### 3.1 Face Localization and Segmentation

As our concern is about the patient's face, we found it wise to constrain our focus on the object of interest (face) rather than the whole image with its background which is of no interest to our study and which does increase the computational burden. There are various methods that bring about the detection of faces in cluttered background, among which our own developed algorithm using Voronoi technique and best described in the work (Dzulkifli et al., 2004) and which is elaborated in details in the journal paper (Cheddad et al., 2004). However, as the mentioned method was meant for grayscale images and as our produced images are in RGB format, we preferred to use the RGB transformation technique highlighted in a bunch of works from which we have chosen the one produced by Hsu et al. (2002). The detection of faces in color images using RGB transformation is simple and yet fast. In the literature there are two main transformation used for this purpose described as follows:

- **RGB ➔ IHS** (Intensity, Hue and Saturation)

Which is a special transformation map called (IHS), which stands for Intensity, Hue and Saturation can be obtained from the RGB bases.

i. Intensity is a measure of brightness:
$I=(R+G+B)/3$ ……1.1

ii. Hue represents the color value:
$H=\cos^{-1}\{[(R-G)+(R-B)]/2[(R-G)^2+(R-B)(G-B)]^{-1/2}\}$ ……1.2

iii. Saturation refers to the depth of the color: $S=1-\min(R,G,B)/I$ ……1.3

In earlier work, Sobottka and Pitas (1996) define a face localization based on (IHS) described earlier, they found that human flesh can be approximation from a sector out of the hexagon depicted in Figure 5 with the constraints: $S_{min}=0.23$, $S_{max}=0.68$,$H_{min}=0^o$ *and* $H_{max}=50^0$.

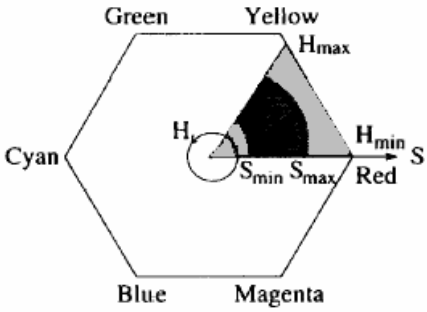


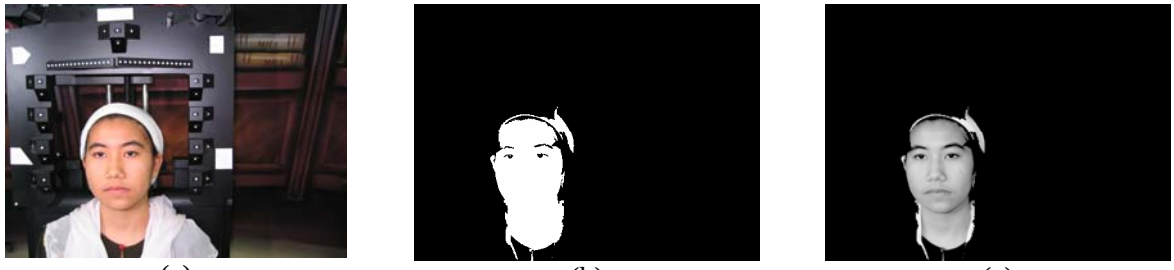

Figure 5: Skin color segmentation in HS space (Sobottka and Pitas, 1996).

- **RGB ➔ $YC_bC_r$** (Yellow, Chromatic blue and Chromatic red)

This is another transformation that belongs to the family of television transmission color spaces, and which is derived from the RGB. Wang and Chang (1997) choose the following system to convert from (RGB) to *($Y$, $C_b$, $C_r$)*:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} (0.299)(0.587)(0.114) \\ (-0.169)(-0.331)(0.500) \\ (0.500)(-0.419)(-0.081) \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad \text{……1.4}$$

A final result of the RGB transformation to yield skin region is shown in Figure 6.



(a) (b) (c)

Figure 6: Face Skin detected using RGB transformation.
(a) Original Color image (b) Face skin region shown as a blob and (c) Intensity color imposed on (b)

### 3.2 Matrix Reloaded

This is not the overheard movie!! This section will discuss how to retrieve back the *fundamental Matrix* (FM) which encapsulates the intrinsic as well as the extrinsic parameters of the stereo system. FM is a compact matrix of 3x3 dimensions that defines and eases the construction of epipolar lines as shown in Figure 7 to solve what is known as the point corresponding problem.

The correspondence problem can be broadly classified into two categories: the *intensity-based matching* and *the feature-based matching* techniques. In the first category, the matching process is applied directly to the intensity profiles of the two images, while in the second, features (corners, circles, lines, edges, arcs, ellipses…etc) are first extracted from the images and the matching process is applied to the features. In our case we choose to deal with the second approach namely features extraction (corners) using Harris Corner detection algorithm. His method of corner detection is based on the gradient or high order derivative. Initially, we need only 8 matched points to construct FM (Harris and Stephens, 1988). Stereo matching process is a very difficult search procedure. Therefore, in order to minimize false matches, some matching constraints must be imposed. Below is a list of the commonly used constraints (Klette et al, 2001):

- **Similarity:**
  Matched features must have similar attribute values.

- **Uniqueness:**
  A given pixel or feature from one image can match *no more than one* pixel or feature from the other image.

- **Ordering:**
  If $m \leftrightarrow m'$ and $n \leftrightarrow n'$, if ***m*** is *to the left* of ***n*** then ***m'*** should also be *to the left* of ***n'*** and vice versa. That is, the ordering of features is preserved across images.

- **Epipolar Line:**
  Given a feature point *m* in the left image, the corresponding feature point *m*' must lie on the corresponding epipolar line.

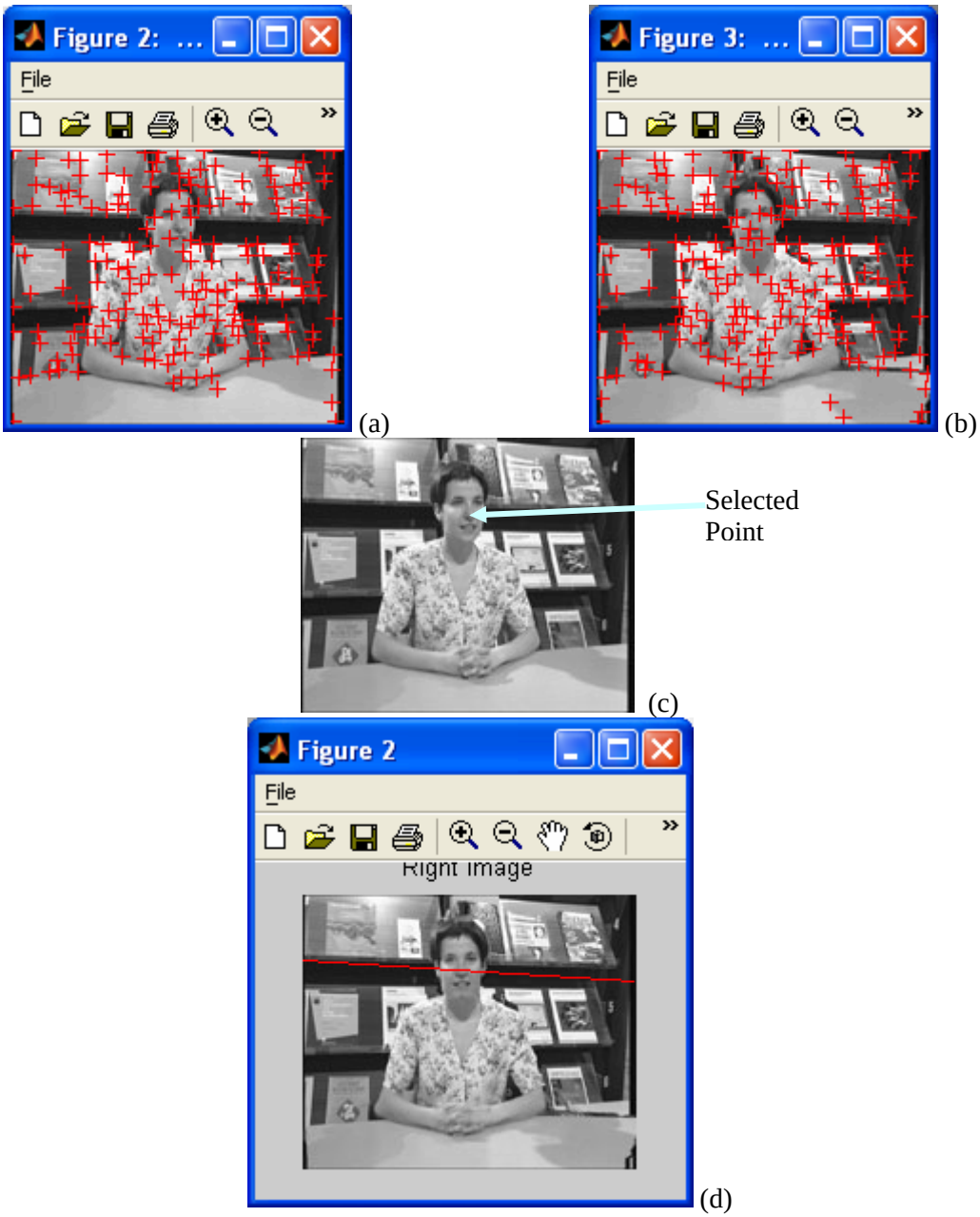



Figure 7: Epipolar line construction. (a) Corners detected in the Left stereo image (b) Corners detected in the right view (c) a manual selected point on the left image (d) The corresponding Point detected on the right image with the help of epipolar line.

Given the 8 matched points (corners) $x_i \leftrightarrow x_i^{'}$ the following system is suggested in the literature to calculate the fundamental matrix:

$$AF = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1 \\ .. & .. & .. & .. & .. & .. & .. & .. & .. \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n'y_n & y_n' & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{bmatrix} = 0 \quad \text{.....1.5}$$

Where
*F* is a nine element vector formed from the rows of F (3x3). A least square solution is applied that minimizes ||*AF*|| subject to the condition ||*F*||=1. This can be accomplished by using the *Singular value decomposition (SVD)*. For instance MATLAB; which is a very powerful Mathematical tool, implements this with the built in function called SVD and which is described as follows:
[U, S, V] = SVD(X) produces a diagonal matrix S, of the same dimension as X and with nonnegative diagonal elements in decreasing order, and unitary matrices U and V so that X = U*S*$V^T$ (MATLAB documentation).

The reconstructed matrix (reloaded matrix) embeds camera intrinsic parameters; which are needed to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame and noted as below:

- Length in effective horizontal pixel size units
- Aspect ratio
- Image center coordinates
- Radial distortion coefficient

After getting this lost matrix (FM), we can easily solve the problem of correspondence in such way that for every pixel in the left image, its corresponding point will lie on the epipolar line which is defined as: $P_R = F * P_L$; where $P_R$ is the corresponding point in the right image and $P_L$ is the selected point in the left image. While the reverse is $P_L = F' * P_R$.

### 3.3 Disparity Map and Depth Calculations

Here we go from the coarse (8 Matched Points) to a fine matching of all points in the target object. In order to reduce the cost of searching into merely 1D vector analysis, we perform; what is known in stereo geometry as, scene rectification, which rotates the position of the epipolar lines to be horizontal ($\theta = 0$) and which is parallel to the base line, this happens physically when the two epipoles are close to infinity (both cameras are facing straight ahead), for more details the reader is referred to the literature (Fusiello et al., 2000). Each line then is examined in the original gray scale right image to locate the optimum response to the correlated sliding window. This is achieved thanks to the normalized cross correlation method, given by the function:

$$\gamma(u,v) = \frac{\sum_{x,y}[f(x,y) - \overline{f}_{u,v}][t(x-u,y-v) - \bar{t}]}{\sqrt{\sum_{x,y}[f(x,y) - \overline{f}_{u,v}]^2 \sum_{x,y}[t(x-u,y-v) - \bar{t}]^2}} \quad \text{.....1.6}$$

Where $\bar{t}$ is the mean of the feature (template) and $\overline{f}_{u,v}$ is the mean of *f(x, y)* in the region under examination.

The above function computes the value of the matching metric using a fixed window (Template) and a shifting window in the targeted image. The shifting window is moved in integer increments along the epipolar line. The optimum score is then selected (Rziza and Aboutajdine, 2001).

When all our matched points are stored in a matrix, we use it as a seed to derive the disparity map as shown in Figure 8 (c), the algorithm is better described in the work of Trucco and Verri (1998). There are different methods (based on how much priori knowledge available about the stereo parameters) that bring about 3D reconstruction. The simplest and most used one is the *Triangulations* (i.e.: to find the minimum distance between the two rays extended from the left point and its corresponding point on the right respectively). Triangulations process can be simplified further if the camera coordinate systems are only translated parallel to each other, which means $Z_L=Z_R=Z$ and in this case the depth (Z) calculation is calculated using some extrinsic parameters to formulate the following equation [Klette, R et al., 2001]:

$$Depth = Z = f * (\frac{T}{Disparity\_Map}) \quad \text{.....1.7}$$

Where, *f* denotes the focal length (in *mm*), *T* is a translation vector and *Disparity _map* ($x_L$ - $x_R$) is the one calculated earlier. A vivid example of the reconstruction view is depicted in figure 9.

(a) (b) (c)

Figure 8: Depth map (a) left image (b) right image and
(c) Depth map visualization where the brighter the pixel is the closer to the camera.

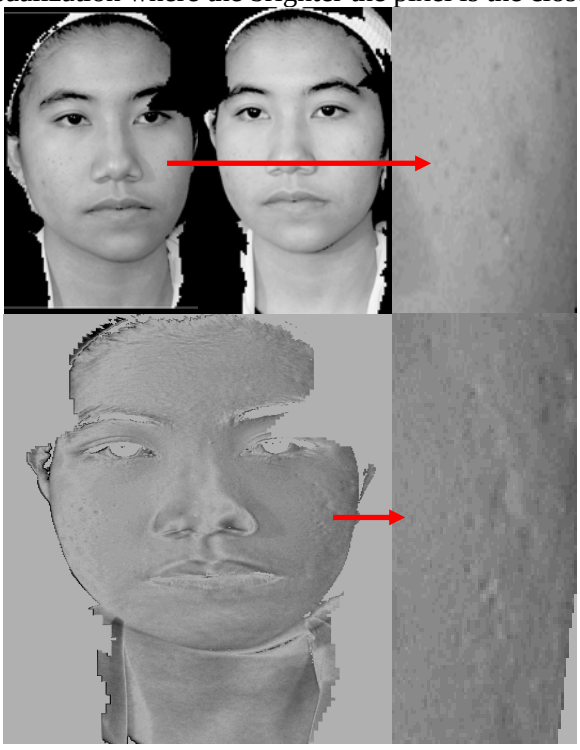


Figure 9: Top: The left and the right image of stereo pair (images were cropped to ease their visualization) with a piece of the check being zoomed. Bottom: The reconstructed image together with the same area zoomed.

The calculated fundamental matrix for the above left and right images is as follows:

F =[ -6.733831e-08  3.460420e-06  -1.057035e-03
  1.430393e-06  1.761693e-07  -1.451869e-03
 -2.171173e-04  -2.429718e-03  9.999954e-01]

## 4.0 CONCLUSION

A precise construction of epipolar lines is of utmost importance in stereo vision since they describe the geometric relationship between the world view points and their projection. In this paper we described how to initiate this construction with merely eight corner points. The object of interest (face) is extracted first using RGB transformation to reduce the cost of computation. The algorithm is being developed for a stereo rig which will be used in conjunction with a project involved in surgery planning for people with abnormal faces.

## REFERENCES

1. Ansari, A-N. and Abdel-Mottaleb, M. (2003). *3D face modeling using two views and a generic face model with application to 3D face recognition.* Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003. 21-22 July 2003 Page(s): 37 – 44.

2. Cheddad, A., Dzulkifli,M., and Azizah, A. (2004). *Exploiting Voronoi Diagram Properties for Face Segmentation and Features Extraction.* Image Processing and Vision Computing, International Journal. Ref.: Ms. No. IMAVIS-D-04-00470.UK (Manuscript under editing).

3. Dzulkifli,M., Cheddad, A., and Azizah, A. (2004). *A New Algorithm for Face Location and Face Features Extraction Based on Voronoi Tessellation and Parametric contour*. IEEE International Conference On Information and Communication Technologies: From Theory To Application. April 19 - 23, Omayyad Palace, Damascus, Syria.

4. Fusiello, A., Trucco, E. and Verri, A. (2000). *A compact algorithm for rectification of stereo pairs.* Machine Vision and Applications. 12 (1): 16 – 22. USA: Springer-Verlag New York, Inc.

5. Halim, S., Zulkepli, M. and Deni, S. (2004). *The Development of Image Capturing System and Information System for Craniofacial Reconstruction*. 3rd FIG Regional Conference Jakarta, Indonesia, October 3-7.

6. Harris, C. and Stephens, M. (1998). *A Combined Corner and Edge Detector*. Proc. Alvey Vision Conf., Univ. Manchester. 147-151.

7. Hartley, R.I. (1995). *In defence of the 8-point algorithm*. Proceedings. IEEE Fifth International Conference on Computer Vision 20-23 Jun 1064-1070. Cambridge, MA, USA.

8. Hassanpour, R. and Atalay, V. (2004). *Delaunay Triangulation based 3D Human Face Modeling from Uncalibrated Images.* Computer Vision and Pattern Recognition Workshop, Conference on 27-02 June. 75 – 75.

9. Hsu, R., Abdel-Mottaleb, M. and Jain, A. (2002). *Face detection in color images.* IEEE Trans on Pattern Analysis and Machine Intelligence. 24(5): 696-706.

10. Klette, R., Schluns, K. and Koshan, A. (2001). *Computer Vision: Three Dimensional Data from Image*. Springer. Singapore.

11. Okutomi, M. and Katayama, Y. (2001). *A simple stereo algorithm to recover precise object boundaries and smooth surfaces*. IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001). 9-10 Dec. 158 – 165.

12. Papadimitriou D. and Dennis T. (1996). *Epipolar Line Estimation and Recitification for Stereo Image Pairs*. IEEE Transactionn on Image Processing. 5 (4): 672-676.

13. Rziza, M. and Aboutajdine, D. (2001). *Dense Disparity Map Estimation Using CUMULANTS*. Figueira da Foz- Portugal 3rd Conference on Telecommunications. 23-24 April

14. Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Publisher: Prentice Hall; 1st edition.

15. Wünstel, M. and Schumann, H.(2002). *Automatic 3D-Reconstruction of the Ocular Fundus from Stereo Images*. CARS 2002 Computer Assisted Radiology and Surgery 16th International Congress and Exhibition June 26-29 Palais des Congrès, Paris, France.