1995

Biometric imaging: three dimensional imaging of the human hand using coded structured lighting

T. A. Vuori

C. L. Smith

Edith Cowan University

Copyright Warning

FACULTY OF SCIENCE, TECHNOLOGY AND ENGINEERING
Australian Institute of Security and Applied Technology

RESEARCH REPORT

Biometric Imaging: Three Dimensional Imaging of the Human Hand using Coded Structured Lighting

T A Vuori, C L Smith

June,1995
ISSN: 1324-6976
ISBN: 0-7298-0195-0

EDITH COWAN UNIVERSITY
PERTH WESTERN AUSTRALIA

Edith Cowan University

Australian Institute of Security and Applied Technology

Research Report 2/95

Biometric Imaging: Three Dimensional Imaging of the Human Hand using Coded Structured Lighting.

T.A. VUORI, C.L. SMITH

Edith Cowan University

Perth, Western Australia

T.Vuori@cowan.edu.au

ABSTRACT

In this report the results of applying a three dimensional range imaging system, based on coded structured light, are presented. This includes a description of a new improved spatial coding scheme. This new scheme increases the number of reference points available and provides a basis for more accurate calculation of their location. A detailed description of the image processing methods used to extract structural information and to identify structural objects from the camera image are given. In addition the method used to calculate the locations of reference points with 'subpixel' accuracy is described. Finally, the results of experiments with synthesised and projected structured light images are presented.

Table of contents

1. INTRODUCTION

A common method of acquiring range data in computer vision is to use structured lighting. The structured lighting is normally implemented with a grid of lased beams or an image projected using a slide projector. If only a grid is provided for the range data extraction an ambiguous situation can be induced by occluded or overlapped images and limit the range data to be only relative. To avoid this scenario a coded structured lighting methods can be used (Posdamer and Altschuler 1982, Boyer and Kak 1987, Le Moigne and Waxman 1988,Vuylsteke and Oosterlink 1990 and Lindsey and Blake 1994). A variation of coded structured lighting requires the combination of the grid and the spatial information in the projected image (Yee and Griffin 1993,1994).

Griffin et. al. (1992) presented a method to generate uniquely encoded light patterns for range data acquisition. Their method used four adjacent neighbours to provide a unique code for each grid intersection. The method of generating these codes is acceptable but the selection of grid primitives for monocrome imaging introduces some practical problems. A new method, where the selection of grid primitives and grid dimensions allows the number of the extracted range points to be four times the number of grid intersections, will be described. This improved method also allows the range points to be evenly spaced and accurately calculated.

2. THE GRID STRUCTURE

To improve the performance of range imaging based on projected grids, a new grid with a new set of grid primitives is proposed. The grid utilises the normally unused areas, the intersections of the lines, in the grid (figure 1). The thickness and spacing of the lines can be selected so that an even distribution of reference point is possible. In this case the corners of squares produced by grid lines are used as reference points. The calculation of unique codes for each intersection is described in detail in Griffin *et. al.* (1992). Only a brief description of how the basic grid and grid primitives are combined follows. In figures 1 to 3 the black areas represent the light pattern projected on the imaged object.

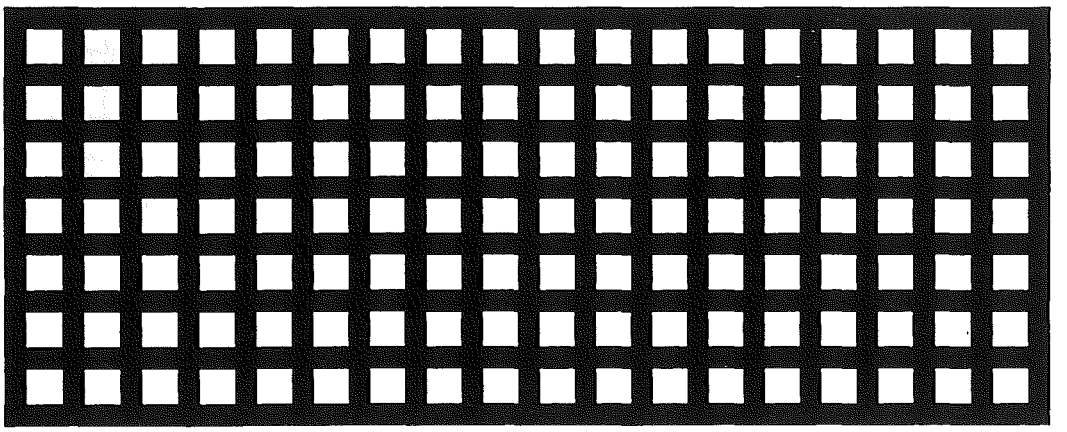

Figure 1: Grid structure

Figure 1 shows a basic grid without any spatial coding. For proposed coding scheme the grid primitives shown in figure 2 are superimposed on the basic grid to produce a uniquely encoded grid. This method of superimposing grid primitives is designed not to interfere with the calculation of the location of reference points in image; the accurate detection of the reference points is the underlying requirement for accuracy in range imaging. In Griffin *et. al.* (1992) only one reference point was produced at each grid intersection. In this new scheme a four reference points are produced. This either reduce the number of grid primitives needed or produces a larger number of reference points with an equal number of primitives.

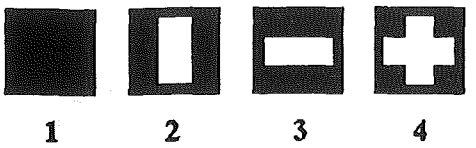

Figure 2: Grid primitives

Figure 3 shows how four basic grid primitives are superimposed on to the intersections of grid lines. The number of grid primitives is purposely kept low to simplify the identification of grid primitives in processed image.

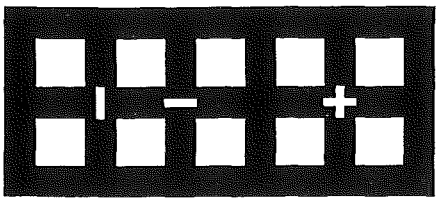

Figure 3: Grid primitives combined with grid structure

After the edges have been detected in the image (figure 4), the corners of each rectangular region can be used for the range points. Each grid intersection will then produce four individual range points. These four range points are connected to one grid intersection which can be identified by decoding the spatial coding embedded into the grid.

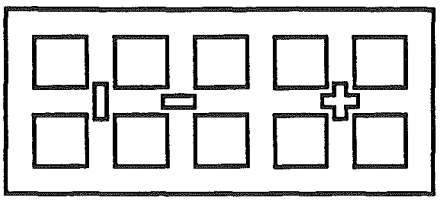

Figure 4: Edge detected image

3. IMAGING SYSTEM

3.1 Imaging setup

Before range data can be acquired, the camera and structured light source must be calibrated. The calibration is performed using the image acquired with the camera. The calibration process involves two steps, the calibration of the camera and the calibration of the structured light source.

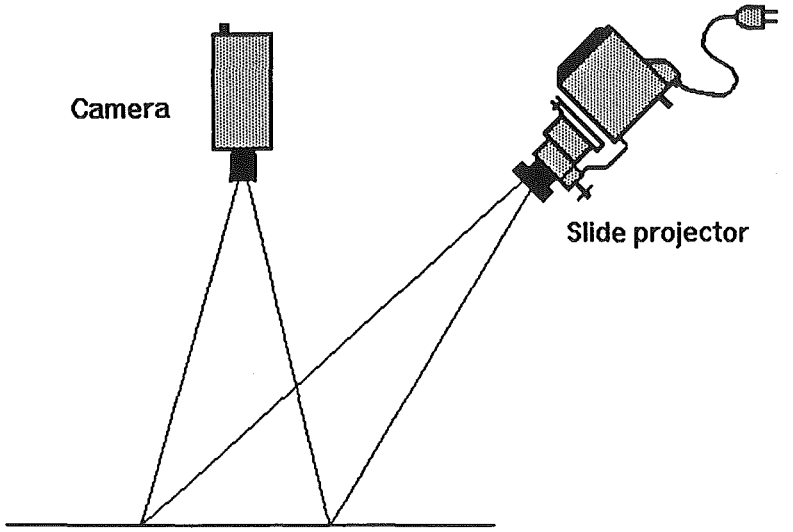

Figure 5: A picture of the imaging system

3.2 Calibration of the imaging system

For accurate triangulation of range data the exact location of the camera and slide projector and their parameters need to be known. In the calibration process the camera is calibrated first. In recent years much effort has gone into developing a fast, accurate and automated calibration method (Tsai 1987, Lenz and Tsai 1988, Weng et al. 1992, Zhang et al. 1993, Wei and Song 1994). One of the successful calibration methods and a detailed survey of other methods is reported by Tsai (1987). Tsai (1987) used a calibration object, with several reference points in known locations, placed in cameras view. This object was used to calibrate camera using an image acquired by camera itself. Similarly in this project the calibration was done by using a known object, a special calibration object, imaged by the camera.

After camera calibration the slide projector was calibrated as it is used to provide structured lighting for triangulation. The first step is to remove the calibration object used for camera calibration. Then project a new set of calibration points into cameras view. This is done with a slide projector using a slide where the location of calibration points is known. By taking into account the properties of camera and it's location, the information acquired with the camera can be used to calculate slide projector parameters.

4. IMAGE PROCESSING

4.1 Edge detection and thresholding

A commonly used method to detect edges in an image is to use a gradient operator. A gradient image is usually produced by applying a mask operator to a whole image. In this study a morphological gradient operator was used to produce a gradient image (Beucher's gradient operator cited in Serra 1982). The output of this operator is simply a half of the difference of min and max values in four pixel neighbourhood. That is :

$$p_1 = (\max(p_1, p_2, p_3, p_4, p_5) - \min(p_1, p_2, p_3, p_4, p_5)) / 2$$

where p_1, p_2, p_3, p_4 and p_5 are the pixel values in four pixel neighbourhood described in figure 6.

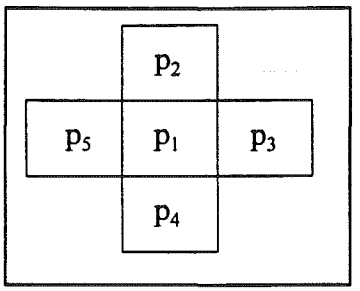

Figure 6: Neighbourhood arrangement used by morphological gradient operator.

The gradient image can then be thresholded. In this case a binary image is produced by using a single threshold T (Gonzalez and Wintz 1987). That is:

$$p_1 = \begin{cases} 1 \text{ if } p_1 > T \\ 0 \text{ if } p_1 \le T \end{cases}$$

The global threshold T was selected based on the grey-level histogram of the gradient image.

An example of the effect produced by the above operations is illustrated below. An image in figure 7 is first processed with the morphological gradient operator and then thresholded to produce the binary image in figure 8.

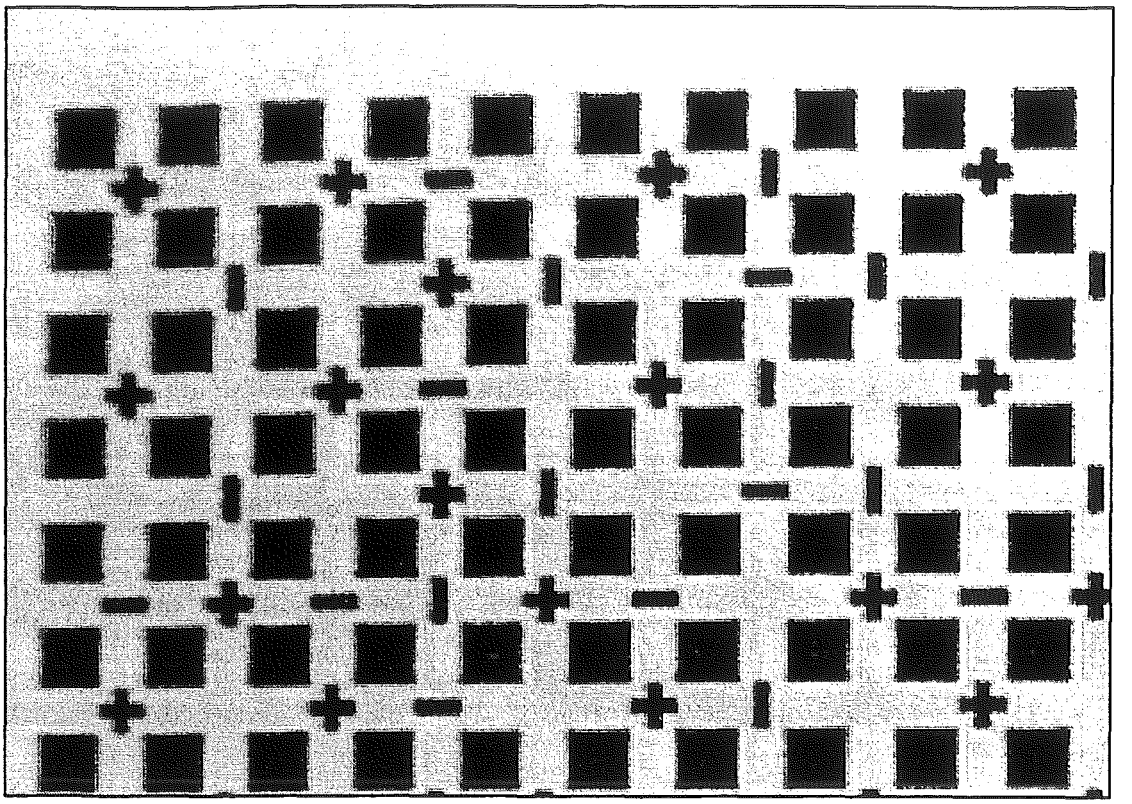

Figure 7: Unprocessed image of the coded grid from camera.

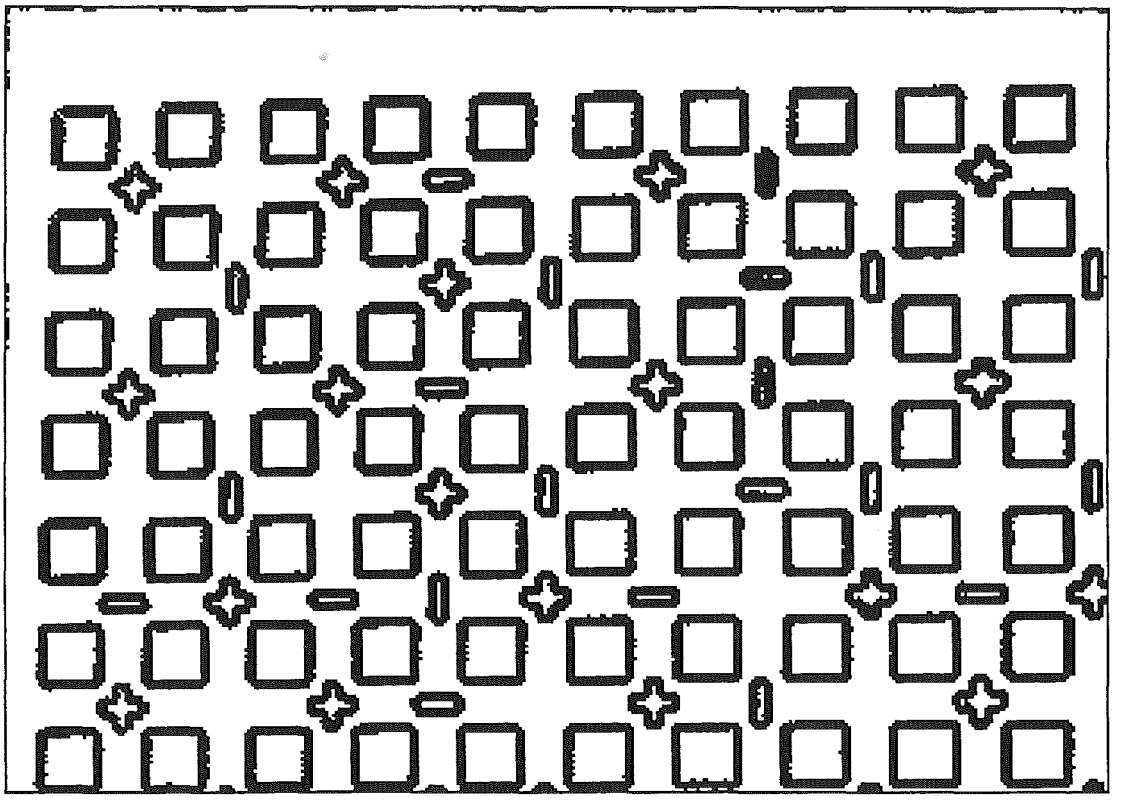

Figure 8: Figure 7 after edge detectection and thresholding.

4.2 Medial axis transform

The result of edge detection is a binary image where value 1 represents an edge region and value 0 an area where no edges were detected (Figure 8). As the interest here is the actual location of the edge, the edge regions need to be reduced to a graph. A practical method to accomplice this is to use thinning algorithm to obtain a skeleton of the region. The algorithm used in this study was originally developed by Zhang and Suen (1984). A good description of the thinning algorithm can also be found in Gonzalez and Wintz (1987) which was used as a basis for the actual implementation. The algorithm is described as follows:

"With reference to the 8-neighborhood definition shown in Figure 9, the first step flags a contour point *p* for deletion if the following conditions are satisfied:

(a) $2 \leq N(p_1) \leq 6$,
(b) $S(p_1) = 1$,
(c) $p_2 \cdot p_4 \cdot p_6 = 0$,
(d) $p_4 \cdot p_6 \cdot p_8 = 0$,

where N(p1) is the number of nonzero neighbours of p1; that is,

$$N(p1) = p_2 + p_3 + \cdots + p_8 + p_9$$

and $S(p_1)$ is the number of 0-1 transitions in the ordered sequence of $p_2, p_3, \cdots, p_8, p_9$.

In second step, conditions (a) and (b) remain the same, but conditions (c) and (d) are changed to

(c) $p_2 \cdot p_4 \cdot p_8 = 0$,
(d) $p_2 \cdot p_6 \cdot p_8 = 0$.

Step 1 is applied to every border pixel in the binary region under consideration. If one or more of the conditions (a) through (d) are violated, the value of the point in question is not changed. If all conditions are satisfied the point is flagged for deletion. It is important to note, however, that the point is not deleted until all border points have been processed."

p_9	p_2	p_3
p_8	p_1	p_4
p_7	p_6	p_5

Figure 9: Neighbourhood arrangement used by the thinning algorithm.

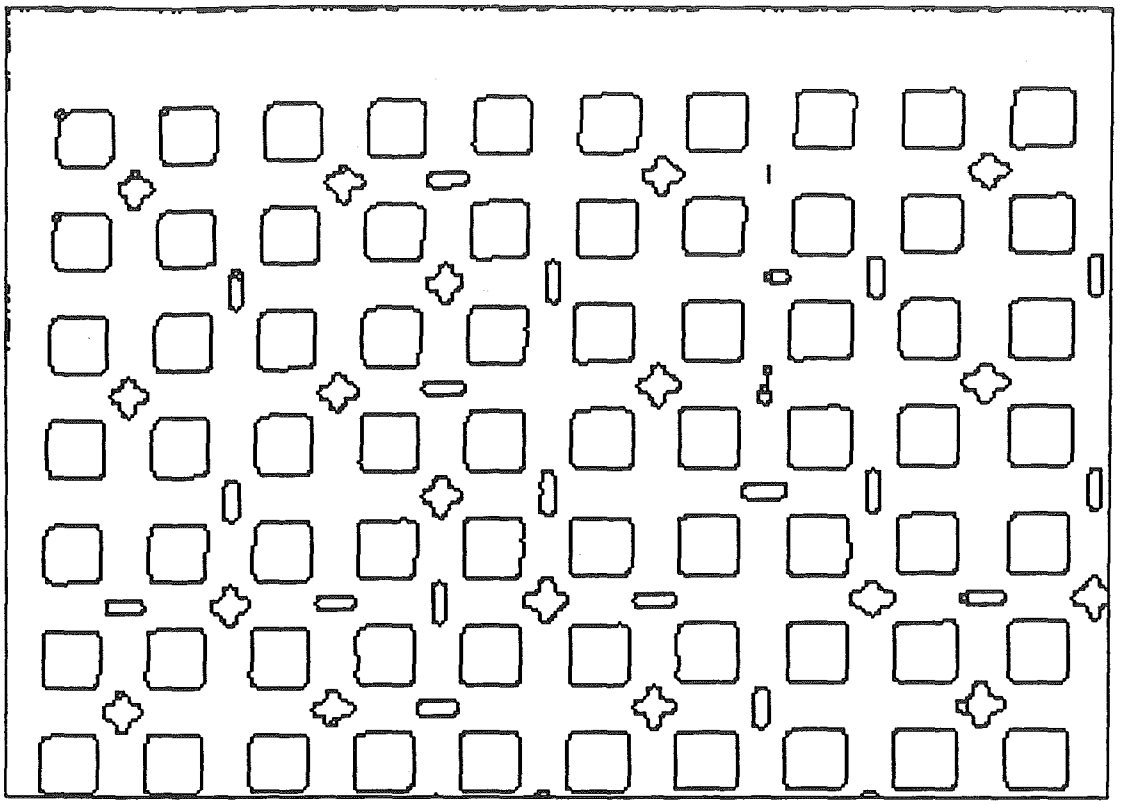

Figure 10: Image after medial axis transformation.

4.3 Object extraction and identification

After medial axis transformation the edges are reduced to single pixels thick graphs. The next step involves tracking all the individual graphs in an image. This is done by scanning the whole image line by line and tracking previously untracked graphs. For each time graph is tracked the object represented by the graph is identified and the type of the object and its location is stored for later use. If the object is a grid primitive, only its type and location is recorded and in the case of the square, its type and corner locations are recorded. If the graph is none of the allowed objects or it starts or ends at the edge of the image, then the object is only partially visible to the camera, it is classified as a noise in image and discarded.

To actually identify the type of each object a simple classification system is used. The classification is based on a few simple measurements of the object. These are the height of object, the width of object, the area enclosed by the objects perimeter and the length of the objects perimeter. To actually identify vertical and horizontal grid objects a ratio of the height to the width is sufficient. Similarly a ratio of the area of object to the perimeter of the object can identify the square object from cross grid object. The use of ratios in classification makes the classification system scale invariant. In other words the size variation of the objects does not affect the object identification process.

4.5 Interpolating corner locations

Figure 11 presents a typical square after extraction. Typically the corners of the square are rounded. This makes the detection of the exact locations of individual corners difficult. As the corners are used as reference points it is important that the actual location of corners is found accurately. To be able to interpolate the location of the corners with "subpixel" accuracy all four sides are described as lines and the intersection of these lines gives the location of each corners. To describe the lines representing sides only two points per line is needed. Optimally these points should be selected along each side as far apart as possible to reduce the digitising effect but not too close to the corners where the rounding is evident. A method where a three pixel long template is fitted as close as possible to each corner for both vertical and horizontal orientations performed well (Figure 11). From each template the closest pixel to each corner was used to interpolate the corner location (see spotted pixels in Figure 11).

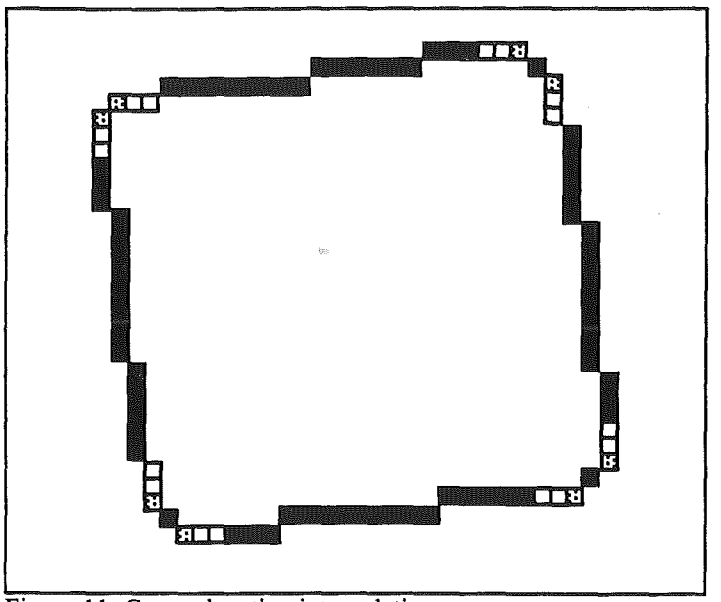

Figure 11: Corner location interpolation.

5. EXPERIMENTAL RESULTS

5.1 The effect of grid resolution on grid object detectability

The resolution of the proposed range imaging system depends on its ability to recognise high resolution grid components. A series of tests were performed to assess the maximum spatial resolution achievable with the proposed grid and using commercial television camera resolutions. Figure 12 shows part of an enlarged processed image. The actual edges are shown in grey and detected objects in black; that is, the calculated corners of the squares are shown as a black dots and the detected grid objects are shown as a vertical or horizontal lines or as a cross. The spatial resolution of reference points, that is, corners of squares, in figure 12 in whole image is 40 x 60. In whole the image that is 2400 reference points in one frame of 768 X 568 pixels. If a smaller projected grid is used, as in figure 13, some of the grid objects are missed and some misclassified. In these experiments the grid was printed using black ink on white paper to achieve maximum contrast. If a projected grid with reduced contrast is used the spatial resolution of reference points needs to be smaller as larger grid object sizes are needed for proper detection.

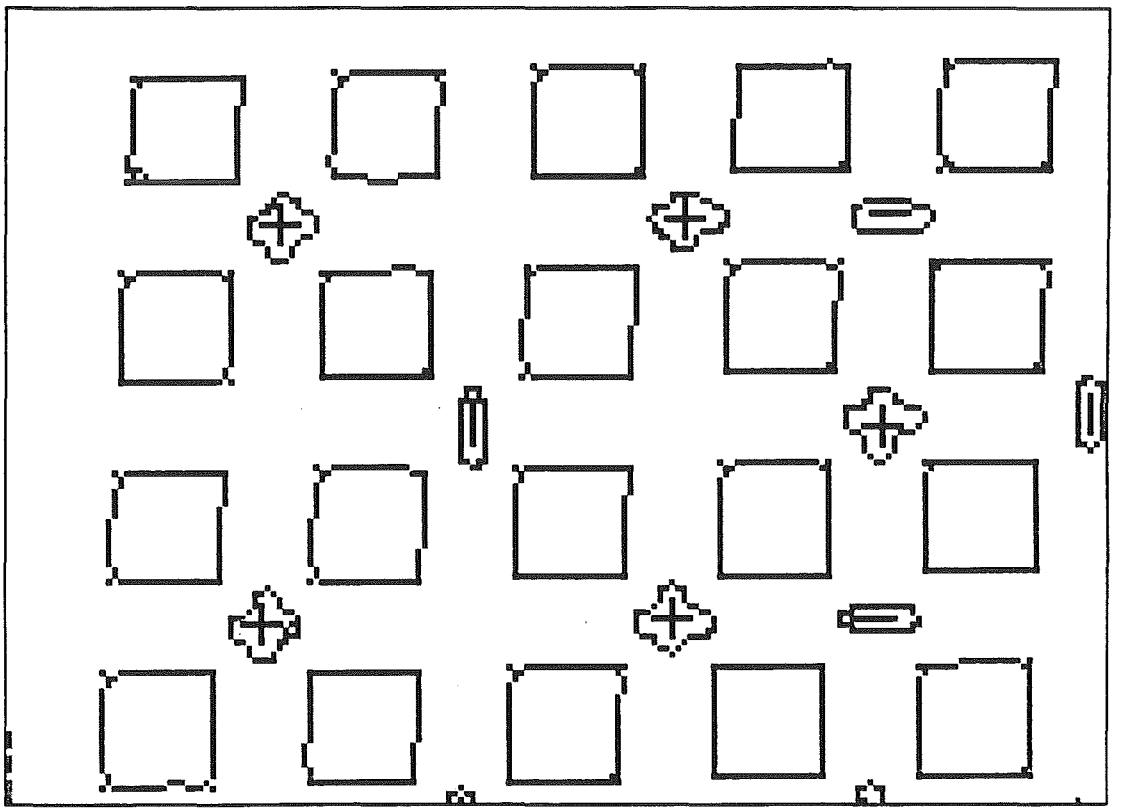

Figure 12: tracked and properly classified image.

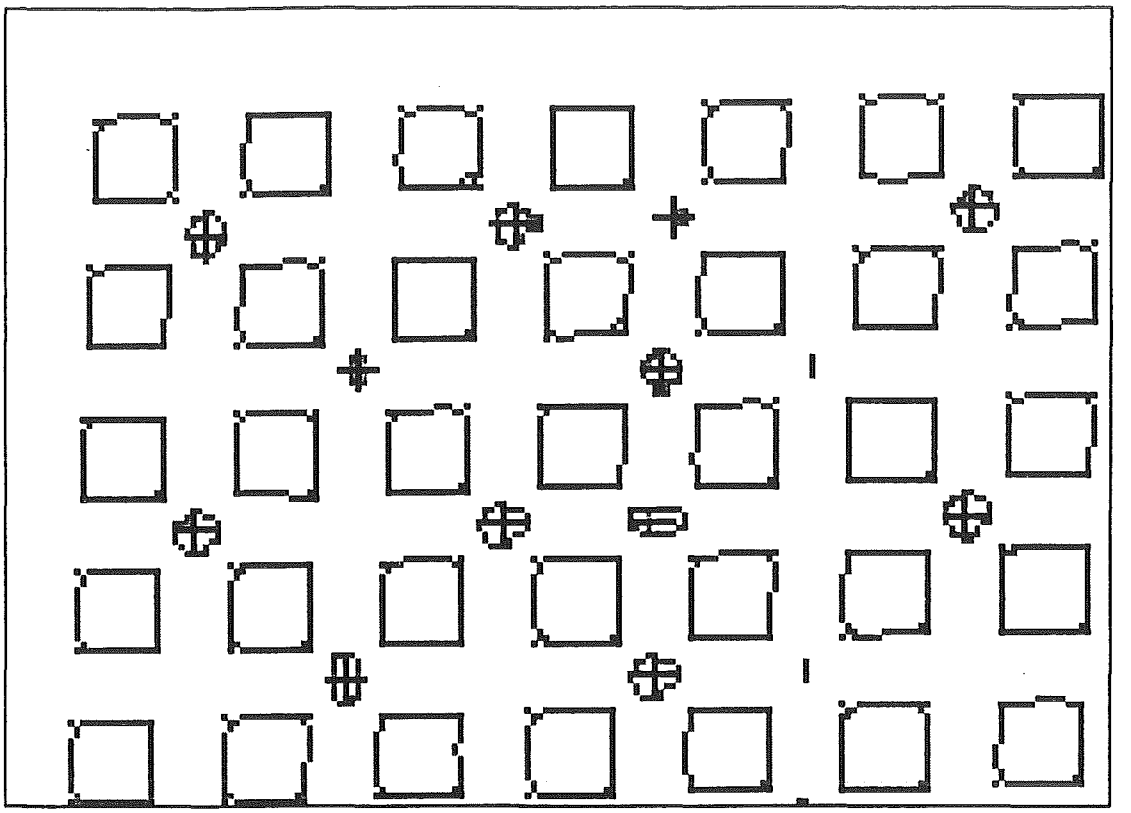

Figure 13: tracked image with erroneously classified objects.

5.2 The effect of surface reflectance on image quality

The basis of range imaging strongly relies on the ability to detect projected grid and grid objects in an actual image. The method used to detect these objects in this study is based on edge detection. The edge detection can only work if the contrast between the projected grid and imaged object is strong. In figure 14 and figure 15 the effect of surface reflectance on image quality is illustrated. In figure 14a an image printed on a paper and imaged with the camera produces a strong contrast between the background and the printed grid. Figure 14b is a projected grid on white paper and figure 14c is a projected grid on Caucasian male hand. As can be seen from figure 14 b and c the contrast between the projected grid and the background is strongly affected by the reflectance properties of the surface. As the surface gets darker the contrast in the image is reduced.

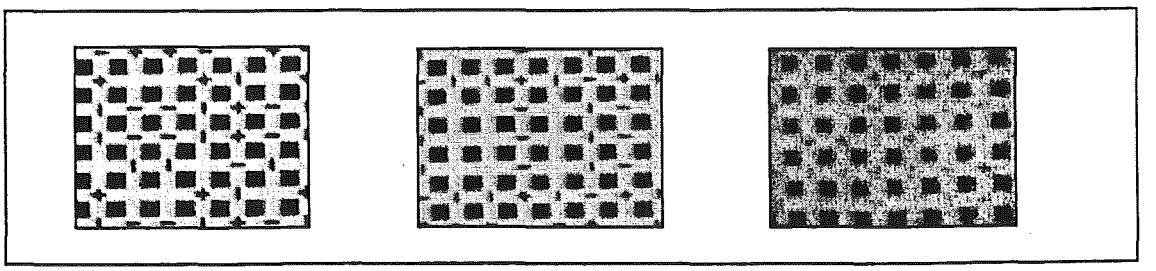

Figure 14: Image quality and surface properties.

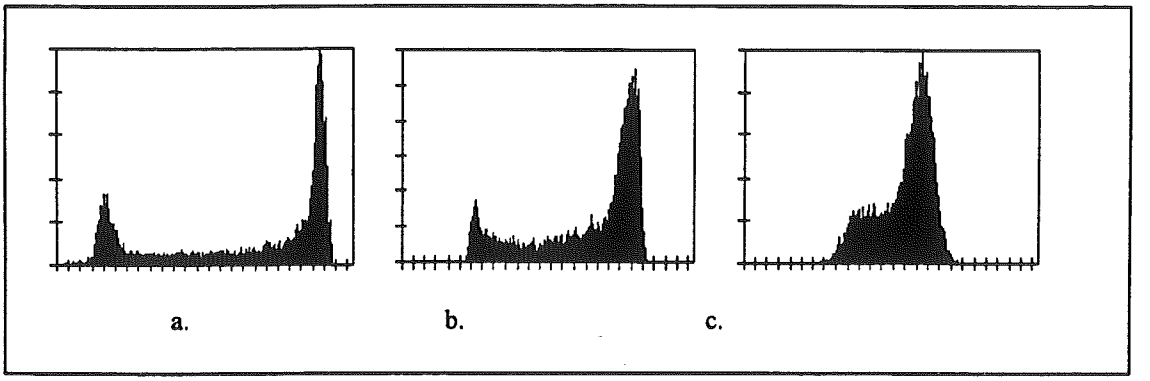

Figure 15: Histograms of images of grid in different surfaces in figure 14.

The reduced contrast is clearly evident in figure 15 were the histograms of images in figure 14 are presented. Note: The measure of variance in pixel values in an image is commonly regarded as the contrast of image. Figure 15a shows a histogram of bimodal image where the actual grid and the background are represented with two peaks. As the contrast in the images in figure 14 reduces, the peaks in their histograms are drawn closer together, until the separation of the two becomes difficult.

5.3 Results with a human hand

The main aim of this research was to produce a range imaging system based on a projected grid for the biometric imaging of a human hand. Due to the nature of the hand the grid needs to be spatially coded to avoid confusion between the reference points in the grid. In figure 15 a grid projected on a human hand is shown. The resolution of the grid in this image is already so high that the grid objects can not be detected reliably. As mentioned earlier, the spatial coding uses the information from five grid intersections to calculate the actual locations. If the fingers in the image are examined and the needs of spatial coding are taken in account, it can be stated that using this method with a commercial television camera resolution, it is impossible to gain proper range information from fingers.

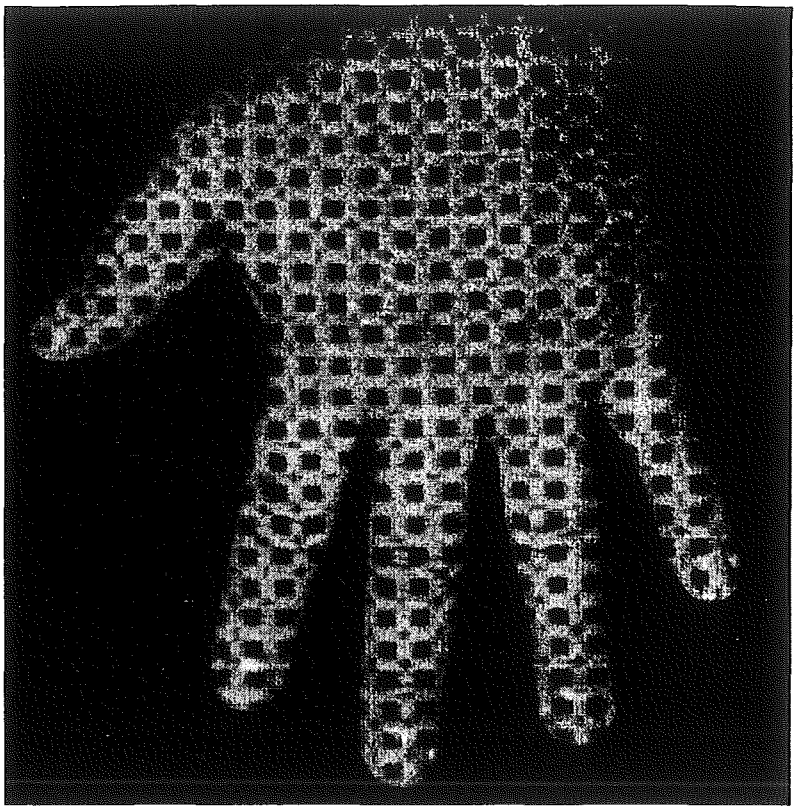

Figure 15: Grid on human hand.

6. CONCLUSIONS

The investigated imaging technique had two serious flaws affecting the practical applicability of the method. The first one is the limited resolution of reference points when commercial television resolution is used in image capture. If higher resolution is necessary the camera resolution can be increased and the number of reference points in image will be increased accordingly. This, though, increases the price of equipment needed to perform imaging and also increases the time needed in processing of the image.

The second difficulty arises when this method was applied to extract a three dimensional image of human hand. The large variations in skin colours and tones sets a tight requirements for both the imaging system and for the edge detection. The underlying problem is a lack of contrast in the projected grid. This makes it difficult if not impossible in some cases to detect the actual grid and grid primitives. In this study all experiments were performed with fairly light coloured skin types and already the lack of contrast was a problem. For darker skin colour the most likeable outcome would be a failure of edge detection to extract the edges of the grid. Also the cracks in finger joints and clusters of hair at the back of the hand produce strong edges in image. These local artefacts in the image will produce extra edges in the image and are easily confused with the actual projected grid. This makes the identification of the grid objects difficult and in some cases impossible.

In light of the above it is recommended that the future directions in biometric hand imaging are directed towards increasing the contrast for the projected grid, removing the grid primitives in the grid and basing the triangulation on the matching of the natural and projected edges from multiple cameras.

7. REFERENCES

Boyer, K.L., Kak, A.C. (1987). Color-Encoded Structured Light for Rapid Active Ranging. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. PAMI-9, No. 1, pp. 14-28.

Griffin, P.M., Narasimhan, L.S., Yee, S.R. (1992). Generation of Uniquely Encoded Light Patterns for Range Data Acquisition. Pattern Recognition, Vol 25, No. 6, pp. 609-616.

Gonzalez, R.C., Wintz, P. (1987). Digital Image Processing. Addison Wesley Publishing Company, pp. 354.

Le Moigne, J.J., Waxman, A.M. (1988). Structured Light Patterns For Robot Mobility. IEEE Journal of Robotics and Automation, Vol. 4, No. 5, pp. 541-548.

Lenz, R.K., Tsai R.Y. (1988). Techniques for Calibration of Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 10, No. 5, pp. 713-720.

Lindsey, P., Blake, A. (1994). Real-time tracking of surfaces with structured light. IEE Colloquium on '3D Imaging and analysis of Depth/Range Images, pp. 5/1-4.

Posdamer, J.L., Altschuler, M.D. (1982). Surface Measurement by Space-encoded Projected Beam. Computer Graphics and Image Processing, Vol 18, pp. 1-17.

Serra, J (1982). Image Analysis and Mathematical Morphology. Academic Press, pp.441.

Tsai, R.Y. (1987). A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. IEEE Journal of Robotics and Automation, Vol. Ra-3, No. 4, pp. 323-344.

Wei, G. De Ma, S (1994). Implicit and Explicit Camera Calibration: Theory and Experiments. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 16, No. 5, pp. 469-480.

Weng, J., Cohen, P., Herniou, M. (1992). Camera Calibration with Distortion Models and Accuracy Evaluation. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 14, No. 10, pp. 965-975.

Vuylsteke, P., Oosterlink, A. (1990). Range Image Acquisition with a Single Binary-Encoded Light Pattern. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 12, No. 2, pp. 148-164.

Yee, S.R., Griffin, P.M. (1993). Range Data Acquisition from an Encoded Structured Light Pattern. Proceedings of SPIE, Intelligent Robots and Computer Vision XII, Vol 2056, pp. 164-173.

Yee, S.R., Griffin, P.M. (1994). Three Dimensional Imaging System. Optical Engineering, Vol. 33, No. 6, pp. 2070-2075.

Zhang, D., Nomura, Y., Fujii, S. (1993). A simple and Accurate camera calibration method. Proceedings of the SPIE - The International Society for Optical Engineering, Vol. 1822, pp.139-148.