3D-NET - A NEW REAL-TIME PHOTOGRAMMETRIC SYSTEM

T. Clarke, L. Li, and X. Wang.
Optical Metrology Centre
City University, United Kingdom
E-mail: t.a.clarke@city.ac.uk

Commission V, Working Group V/1

KEY WORDS: real-time, 3-D, photogrammetry, metrology, digital signal processor

ABSTRACT

The move from computer/ frame-grabber/ camera combinations to networked cameras with onboard processing is progressing rapidly. This development is long overdue and will produce a significant change in the way in which embedded close-range photogrammetric systems operate and what they are capable of. It will become feasible to track the 3-D position of multiple objects over large areas with high accuracy and reliability. This will be increasingly important for applications such as: virtual reality environments, tracking surgical instruments during surgery, or monitoring assembly processes in the manufacturing environment. This paper describes the development of a number of intelligent camera nodes designed for photogrammetric measurement purposes. Each node consists of a video processor board, which performs real-time extraction of targets locations from images and a digital signal processor which recognises targets and calculates their sub-pixel locations. The target locations are then transferred to a host computer for 3-D estimation. Each camera system is capable of producing 2-D estimations of target image locations at a sustained rate of over 170 targets every 1/25 of a second.

1. INTRODUCTION

A programme of development of a network based real-time measurement system began at City University in 1994. Some initial results were published (Pushpakumara, 1995; Gooch, et al, 1996(a & b); Pushpakumara et al, 1996; Wang & Clarke, 1996;) concerning this work and an overview paper was presented (Clarke, et al., 1997). This paper discusses the ongoing development of this system that uses a number of networked intelligent cameras.

2. 2-D PROCESSING

2.1 Hardware

The 2-D processing hardware is based on the Analog Devices ADSP-21xx family of processor. This modular system consists of a DSP module (DSP-90), a general I/O (GPIO-90) module, a video feature extractor (VFE-90) module, an Ethernet communications (ETH-90) module and a power supply unit (PSU-90) module. Each camera contains an embedded DSP-90 system where images of retro-reflective targets are processed and sub-pixel 2-D co-ordinates of the targets are calculated (Figure 1).

The VFE-90 module is a hybrid circuit, comprising both analogue and digital circuitry. By performing processing at hardware level the data requiring processing is reduced considerably. This makes it possible to achieve real-time photogrammetry at a reasonable cost. After processing by the VFE-90 module, only the line-by-line video signal (A-D converted into 16 bits words) which is above the threshold level is stored in a First In First Out (FIFO) buffer. If there is no object above the threshold a value denoting the end pixel location is placed into the FIFO. For a line with a target a pixel location together with the intensity of the first edge along with all subsequent contiguous pixel intensity of each target image are also stored in the FIFO. A bit flagging the beginning of each new frame is encoded into the pixel location word (The intensity is a 10-bit quantity leaving 6 bits free for other uses). For interlaced imagery the odd-even field output of the synchronisation stripper is used to direct the data to one of two FIFO's (A and B), one for odd line data and the other for even line data. For a camera which is imaging a number of targets evenly distributed throughout the image the FIFO's will be filled in the following way. The FIFO's are reset; this has the effect of emptying them. Data is read from the FIFO's until data starts going into FIFO B. A new frame can be extracted at the point when FIFO B has been filled with one field's data and FIFO A is just filling up. Data from FIFO's A and B must be combined to produce image data corresponding to a frame. Odd and even lines can be taken from both the FIFO's by reading them alternately. This means that there is a delay of 1/50 of a second before processing can begin on the frame (figure 2).

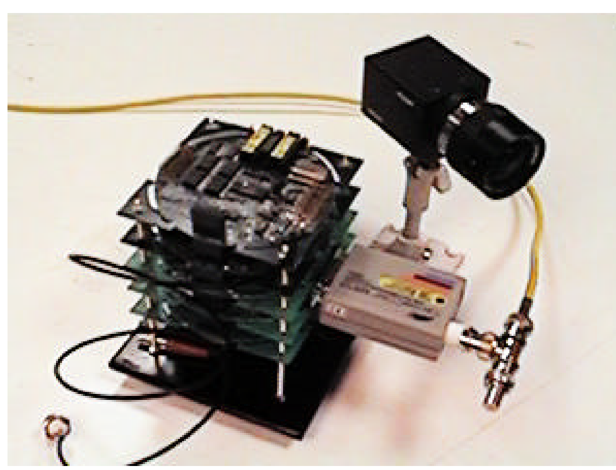

Figure 1. Image of DSP-90 networked camera system.

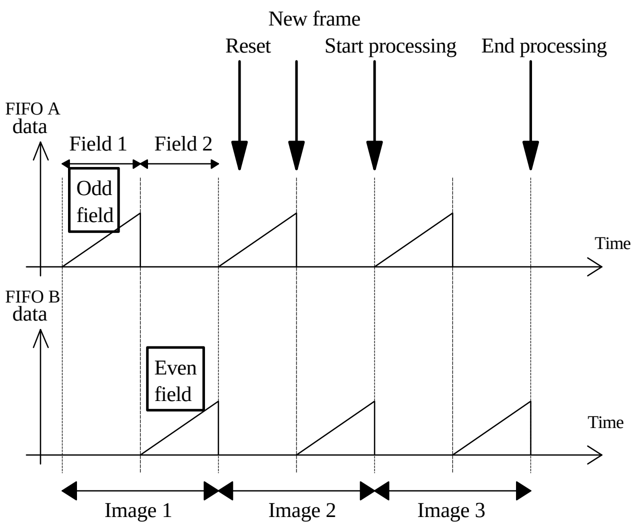

Figure 2 Graphical representation of filling the FIFO's with data and frame alignment after reset.

2.2 Image processing

2.2.1. Target image processing. Image acquisition is a fundamental process for photogrammetry. Prior to obtaining any video data, the VFE-90 module must be reset. For an interlaced imaginary a new frame and subsequent frames must be aligned. Then, the video signals are processed line by line. Three types of data: sub-pixel location, intensity, or new frame, are read out from the FIFO's. Decoding the data gives meaningful edge and intensity information of target sections. Figure 3 illustrates extracted data from the FIFO's with a threshold set at 50 for three targets (two of which overlap each other in the same line) in an interline image. The threshold value is set by an 8-bit I/O port on the GPIO-90 board. The lower two bits of the intensity are not used so a threshold of 50 equates to 200. At this stage in the development of the *3D-NET* system the video signals have not yet been adjusted optimally.

```
Line_1        eoln
Line_2        eoln
.
.
.
Line_377  eoln
Line_378  eoln
Line_379  edge 354 267  302  286  271  252  205  eoln
Line_380  edge 353 244  335  385  385  349  305  243  eoln
Line_381  edge 353 249  339  395  397  355  315  248  eoln
Line_382  edge 353 214  279  317  306  280  260  209  eoln
Line_383  edge 355 216  211  eoln
Line_384  eoln
Line_385  eoln
Line_386  eoln
Line_387  edge 346 272  331  354  326  295  256  eoln
Line_388  edge 346 219  254  261  246  229  206  eoln
Line_389  edge 346  254  300  309  294  264  227  edge 356 254  278  265  245  226  eoln
Line_390  edge 345 209  295  358  370  349  308  256 edge 355 250  332  366  348  324  282  214  eoln
Line_391  edge 347 210  212  204 edge 355 246  325  359  342  317  275  210  eoln
Line_392  edge 355 221  269  287  279  255  228  eoln
Line_393  edge 357 218  213  210  eoln
Line_394  eoln
.
.
.
Line_ 581       eoln
Line_582        eoln
```

eoln = end of line marker
edge = beginning of a new line object
Line number are calculated by counting eoln markers

Figure 3. Data Extracted from FIFO's A and B

The target location algorithm only requires edge data belonging to two consecutive lines of an image at a time to compute the sub-pixel location of each object. A parameter buffer is used to store the peak intensity, a summation of intensity and summation of **x** or **y** location times intensity for each object. The data for consecutive lines are stored in two buffers "ping" and "pong". The ***ping*** buffer of current line becomes the ***pong*** buffer for next line target recognition. The edge pairs in the ***ping*** buffer are compared with those in the ***pong*** buffer to ascertain the state of the edge pairs in the current and previous lines. By comparing the starting and finishing pixel values of target sections in each two consecutive lines, the targets of legitimate shapes present in the frame are reconstructed. Splitting or merging targets are recognised and flagged as invalid photogrammetric target images. The sections of invalid targets within subsequent lines are processed but no sub-pixel location is computed. The completely reconstructed targets are assessed for validity using the area and peak parameters.

2.2.2. DSP programming issues. The ADSP-21xx family base architecture provides single-cycle computation for multiplication together with accumulation and supports extended sums-of-products. Detecting target edges and accumulating grey scale intensity value for each target can be achieved efficiently with a DSP. These data are used to calculate the grey-scale centroid of target.

2.3 Image location

2.3.1. On-the-fly centroid computation. The main task of the image location algorithm is to calculate the grey scale centroid of the targets. Object of all shapes have to be processed but only the objects that meet the criteria set by the target image recognition algorithm have to be located accurately. One method of computing the centroid would be to create sub-images (together with an offset from the origin) for conventional processing. This would involve storage of the image in a temporary location and extra computations. The DSP has limited resources for such tasks and therefore another method was chosen. This method uses the line-by-line approach used by the image recognition algorithm to compute the centroids at the same time as the objects are being recognised. In this way the storage of information is limited to the data for the summations required for each of the objects. The DSP is highly efficient at multiply and accumulate operations required in the centroid calculation. However, the DSP used is a fixed-point device which multiplier produces a 32-bit product. When the accumulation has overflowed beyond the 32-bit boundary the object is unlikely to be a retro-reflective targets and so the accumulation stops, but the algorithm must still deal with the object to avoid problems in subsequent lines of the image.

2.3.2. Criteria for selection of legitimate targets. The *3D-NET* system was designed with the use of retro-reflective targets in mind, these targets produce images of predictable size and shape. Other features above the threshold will often be of a different size and shape. To recognise target-like features one or two measures can be used on-line to select candidate targets. Two intuitive and easy to implement measures, which nevertheless give most of the information required, are the area and peak intensity. The area is simple to accumulate on-line being the sum of line lengths for each object, the peak requires the comparison of current intensity with the previous intensity and the storage of the greatest. Other measures such as radius or maximum differences of **x** and **y** co-ordinates could be used but are likely to provide some redundancy of information.

2.4 Results

A series of tests were conducted to assess the performance of the intelligent camera system. The processing time is dependent on the image data. The aperture of the camera was used to vary the image size from a minimum size to a maximum where the targets were saturated.

The first test evaluated the maximum number of targets to cause an overflow of FIFO A which has to contain the field data until processing can begin. The average area of the targets was computed for various settings and the results are given in Table 1.

Average Area (pixels)	Number of targets
39	232
34	255
32	274
30	287
29	295
27	309
21	340

Table 1. The maximum number of targets capable of being processed without overflowing the FIFO's.

In a typical scene of retro-reflective targets not all targets will be oriented optimally towards the camera and will thus be relatively dim compared to others that are in optimal orientations. As a result an average area of targets of some 20 - 30 pixels would allow high precision measurement of some 280 to 340 pixels using the each intelligent camera system. This figure is independent of the processing speed and is purely a function of storage capacity of target images using the image-encoding scheme. It is worthy of noting that some 300 targets can be stored in around 16k words of DSP memory. This represents a considerable reduction in storage space and threshold processing of these 430k images. This figure would be the same regardless of whether 1kx1k or 3kx2k images were used if the target numbers and image sizes were the same.

A second series of tests were conducted to assess the real-time capability of the system. In this case some 100 frames were processed to ascertain the number of targets required to produce a FIFO overflow. An overflow of the FIFO's shows that the DSP is incapable of keeping up with the processing task. To make this test realistic the Ethernet communications stage was also used so the sub-pixel target locations together with the area and peak of each target were transferred. Table 2 illustrates the results.

Area of target (pixels)	Number of targets
49	107
38	132
35	148
31	170

Table 2. Real-time 2-D target image processing capability.

The results show that for typical target sizes with a diameter of some 6 - 7 pixels (optimum of target image location) some 170 target images could be processed in real-time and the results recorded by another computer. These result were conducted using the DSP with a 12 MHz clock instead of its designed 20 MHz clock speed because of a timing problem that was being experienced at the time. Use of the DSP at 20 MHz would result in approximately 1.7 times more targets being processed in real-time. Preliminary tests on the resolution of the target location algorithm for multiple frames revealed the expected variability for a stationary target (smaller than 0.01 of a pixel).

3. NETWORK COMMUNICATIONS

3.1 Ethernet hardware

The interface between a DSP-90 system and a PC are implemented via the ETH-90 Ethernet communication module. Ethernet is a fast means of communications with a number of useful features such as 10 Mbits/second data rate (approx. 1 Mbyte per second), up to 230 metre cable length, multiple clients and servers, packet collision detection, software support, etc. The ETH-90 module contains a National Semiconductor's Serial Network Interface Controller (SNIC) solution providing Media Access Control (MAC) and Encode-Decode (ENDEC) functions in accordance with the IEEE 802.3 standard. Its interrupt signal is connected to the DSP-90 module.

Windows Sockets provide a means of fast data transfer into a PC program running using various protocols. A programmer can create a standard windows application incorporating the full functionality of a network without knowing anything about a specific Ethernet card that is being used.

3.2 Ethernet software

The ETH-90 module must be initialised prior to transmission or reception of packets from network. After the initialisation the module senses when data arrives and interrupts the DSP processor so that the data can be collected. A simple message-passing scheme is used in the *3D-NET* system for control of processes. At the beginning of each packet data (excluding headers) a control word is used to indicate the process that is required. The DSP-90 module program has a loop in which calls to various sub-routines are made depending on the control word. Both sides of the communication, PC and DSP, know exactly what to do following a given command. For instance, a command may be a request for data in which case the PC will expect data of a certain type to be returned, in the case of a set up command a confirmation of correct operation message may be returned.

Various methods are defined in the TCP/IP protocol and are implemented as a basis for low or high level communication between Ethernet nodes. For instance: methods of establishing: the hardware identity of other nodes (ARP), whether other nodes are operating (PING), or reliable connections between peers are defined. In this implementation IP datagrams were selected as the best method having low overhead and the capability of performing fragmentation and de-fragmentation of large packets. Datagrams are referred in the literature as *unreliable* as there are no checks for successful reception of data. However, errors in the sent data are unlikely. By implementing a one-to-one scheme for the DSP-90 nodes it is possible to avoid most of the problems that require, so called, *reliable* connections.

The Ethernet standard defines that a packet can vary between 46 and 1500 bytes in length. If the packet is a shorter than 46 bytes the software is responsible for padding or stripping pad bytes. For a packet larger than 1500 bytes the packet must be fragmented or reassembled.

3.3 Protocol

A standard IEEE 802.3 packet consists of the following fields: preamble, Start of Frame Delimiter (SFD), destination address, source address, length/type, data, and Frame Check Sequence (FCS). All fields are of fixed length except for the data field. To achieve Ethernet communication between ETH-90 module and a Window socket program an Internet Protocol (IP) header is used within data field. The 20 byte header is a small price to pay with a large packet and for a small message packet the header is within the 46 byte minimum size so represents no problem.

3.4 Results

To assess the overhead of communications using the ETH-90 module a test was conducted to ascertain the number of targets that could be processed with and without the communications. The results are given in table 3.

Number of targets with comms	Number of targets without comms	Overhead (%)
206	219	5.9
166	176	5.6

Table 3. Communications overhead results.

Once the Ethernet chip has been set up there are relatively few tasks required to send or receive packets. The main one is the placing or collecting of the data into the memory of the ETH-90 module. The sending or reception of data takes place independently of the DSP-90 operations. The 6% overhead represents some 2-3 mSec of time. This could be reduced further by software improvements. The Windows Sockets implementation is able to communicate with any reasonable number of DSP-90 systems by use of unique IP addresses for each system.

4. 3-D PROCESSING

The developers of the *3D-NET* system have sought to provide a system with ideal characteristics for real-time operation. One of the most desirable characteristics of such systems is predictability of the delivery of information e.g. 3-D co-ordinates. This priority has dictated research into the computational nature of solving correspondences and computing 3-D co-ordinates of the corresponding targets.

4.1 Correspondence

Calculating the correspondence between target images can be computationally expensive. While some methods that use stereo matching between cameras followed by back projection produce a working solution it cannot be considered ideal for real-time systems. This is because targets viewed by cameras 3, 4, and 5 cannot be corresponded, and the collinearity equations are used which are more expensive than 2-D calculations. An approach which is better suited for real-time operations is rectification (figure 4) and combination searching as this reduces the search complexity from a 2-D search for correspondences or ambiguities to a 1-D search (Ariyawansa & Clarke, 1997). This is exactly the required predictable characteristic.

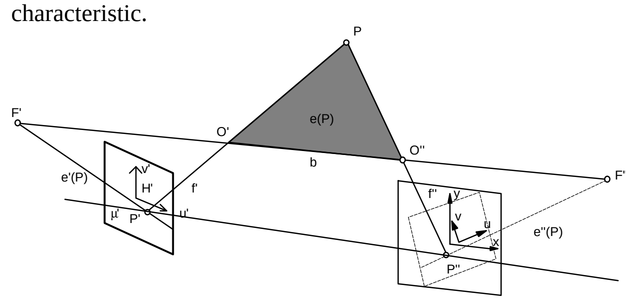

Figure 4. Rectification of convergent images.

4.2 Separated adjustment for real-time applications

Least squares estimation has been successfully used for redundant measurements in close range photogrammetry. Conventionally all the unknown parameters, the 3-D co-ordinates of targeted object points and the camera parameters, are estimated simultaneously. This leads to the bundle adjustment, which is expensive in terms of computation time and memory requirements making it unlikely to be used in real-time applications. To solve the problem an alternative method called the separate adjustment has been developed by Wang (Wang & Clarke, 1996 and 1998) and tested in this real-time photogrammetric system. The idea of this method is to estimate the 3-D co-ordinates of object points and the camera parameters separately and iteratively. Due to the low computation requirements and the linear computational complexity, the speed of solution of the separate adjustment is much faster than that of the bundle adjustment. The maximum memory required by the separate adjustment is limited to a 6×6 unit no matter how many cameras and object points are involved. For a four-camera system 3-D co-ordinates of one hundred object points can be calculated by the separate adjustment method using standard desktop computer.

4.3 Results

To assess the real-time capability of the correspondence and 3-D estimation some tests were performed. A 133 MHz Pentium Computer was used running Microsoft Windows 95. A correspondence algorithm and the 3-D estimation process were tested for varying numbers of targets and the time estimated for each operation (Table 4).

No of targets	Correspondence (mSec)	3-D estimation (mSec)
28	4	8
53	7	15
81	11	23
143	18	39

Table 4. Time for correspondence and 3-D estimation.

These times show that real-time processing is a possibility given that the image-processing task takes place in parallel and picking up the 2-D data by the PC is likely to be very quick. A faster processor could also be used to provide several times more processing power if required.

5. APPLICATIONS IN JIGLESS ASSEMBLY

5.1 Jigless manufacture

The *3D-NET* system is expected to find uses in jigless assembly. In the past the Aerospace industry has relied heavily upon jigs to produce assemblies of components to tight specifications. All of the key information is tied up in the jig and as a result these have to be measured to ensure that they are within specification (figure 5 illustrates the current methods of assembly of Airbus wings). It is currently perceived that large static jigs constrain the manufacturing process in an undesirable way. They delay the time to market of new products, fix the method of manufacture for the lifetime of the product (up to 30 years for a civil aeroplane wing), and are expensive to make and periodically certify. Civil Aviation Authorities can require the aerospace industry to certify its jigs. Some jigs have never been measured and in some cases no CAD design exists. In the past errors in the manufacturing processes have resulted in large qualities of shims which add the weight of an aircraft and ultimately to the cost of flying that craft via the extra fuel required. For instance, a new version of the Hercules aircraft has resulted in a 40 percent greater range, cruising ceiling and decrease in take-off distance together with a 21 percent increase in maximum speed and a 50 percent decrease in time to climb (Brown & Sharpe, 1998). Measurement can play a vital role in this process by reducing unwanted shimming and product material.

The move towards jigless assembly or what may also be called minimal tooling assembly (as some tooling will always be required) will take place using a mixture of humans, actuators, and robots (figure 6). The objectives of the "jigless" method are the: reduction of manufacturing costs; provision of greater manufacturing flexibility; avoidance of expensive certification of jigs; faster product to market times; and improvements in the speed of manufacture. Measurement systems provide essential information to ensure correct manufacture. In the future the jigging function will be provided directly from CAD via the measurement function rather than indirectly as at present.

Figure 5. Airbus Wing
(Picture courtesy of British Aerospace)

A photogrammetric systems best advantage in this environments is the ability to simultaneously measure many points at one instant the disadvantage being the necessessity for retro-reflective target to be used at each point to obtain the highest accuracy.

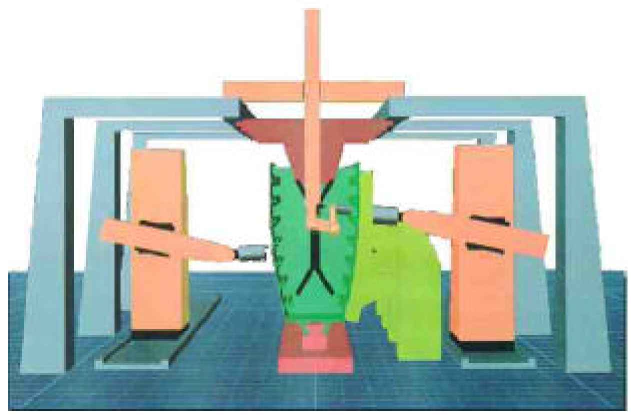

Figure 6. Advanced assembly cell visualisation
(Arrow, 1998)

5.2 Part positioning task

To start the process of embedding the *3D-NET* system in a manufacturing cell a demonstrator was produced to illustrate the capabilities of the system and allow development of techniques. The demonstrator (figure 7) consisted of two parts each with a five degrees of freedom movement. One part had a laser mounted on it and the other had two small pinholes. The task of the measurement system was to measure the relative position of the two parts such that the laser would pass through the two pinholes. One of the two objects could then be moved to an unknown position and the transformations necessary were calculated to move the other object back into the same orientation. This process required each object to be recognised and tracked.

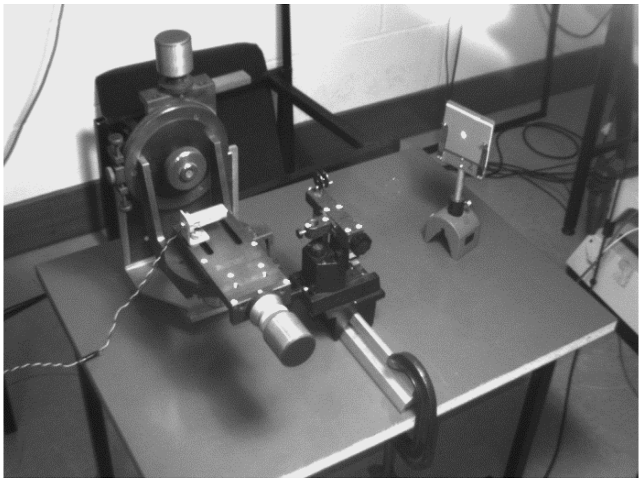

Figure 7. Image of demonstration system.

5.3 Software

The 3-D co-ordinates of the object points obtained from the separate adjustment or the bundle adjustment are in an arbitrary datum if control points are not involved. It is necessary to transform these object points from one co-ordinate system (the arbitrary co-ordinate system) to another co-ordinate system (a given or common co-ordinate system) for the purpose of relative positioning. The relative position **X,Y,Z** and rotation α,β,γ can then be estimated. Control points can be used to define a common co-ordinate system.

5.4 Results

The part positioning demonstration used four Pulnix TM 6CN cameras that were pre-calibrated to a precision of 0.2 μm rms (approx 1/40 of a pixel). The cameras were set up with a convergent angle of 60 degrees between the camera and the normal to the laboratory floor. The camera to object distance was 2.2 metres and the focal length of the lenses was 16 mm. The resulting 3-D precision of the object space co-ordinates was estimated to be 20 μm rms.

At the time of writing a three-camera system has been put together and the processing units have been coupled to the measurement cell. The cameras are working as expected and the time to repeatedly collect 2-D image data, compute correspondences and produce 3-D co-ordinates is about $1/10^{th}$ of a second. Most of this time is taken up with resetting the FIFO's for each image. Further cameras will be added to the system and the current software will be modified to take into account the requirements of real-time systems.

6. CONCLUSIONS

This paper has discussed the on-going development of a real-time close range photogrammetric system – *3D-NET*. The requirement for real-time measurement in aerospace applications has resulted in a system which uses a number of features which are novel that when put together produce a unique system. 3-D measurement at a continuous rate of 100-200 targets at camera frame rate together with the flexibility of a networked system is an achievable goal. Future work will apply this system to real-time jigless assembly tasks. An upgrade to the Analog Devices 2101 processor with a Analog Devices SHARC processor which is floating point and runs at 40 MHz is expected in the next six months. This is predicted to more than double the performance of the system making real-time measurement of 300 points a possibility. Other extensions of the system to digital cameras may also be contemplated if there is a demand.

7. ACKNOWLEDGEMENTS

The work of D.D.A.P. Ariyawansa, and Dr. R.M. Gooch is acknowledged in the development of the *3D-NET* system.

8. REFERENCES

Ariyawansa, D.D.A.P, and Clarke, T.A. 1997. High speed correspondence for object recognition and tracking. SPIE Vol 3174, Videometrics V, pp. 70-79.

Arrow. 1997, BAe magazine, Issue 13, December, p 11.

Brown and Sharpe, 1998. The World's Largest Planes Get A Lift From Co-ordinate Metrology, MFG The Brown and Sharpe Publication of Precision Manufacturing, Winter 1998, Vol 5, Issue 1, pp. 4-7.

Clarke, T. A. Gooch, R. M. Ariyawansa, D. D. A. P. Wang. X. 1997. *3D-NET* - the development of a new real-time photogrammetric system. Videometrics, SPIE Vol 3174. pp. 222-233

Gooch, R.M. Clarke, T.A. & Ellis, T.J., 1996. A semi-autonomous sewer surveillance and inspection vehicle. Proc. IEEE Intelligent Vehicles Symposium, Japan, pp 64-69.

Gooch, R.M. Clarke, Pushpakumara, D.D.A., & Ellis, T.J., 1996. Proc. DSP 96 Scandanavia. Copenhagen, Denmark, pp. 159-168.

Pushpakumara, D.D.A. The Use of a Digital Signal Processor in real-time image processing applications, MSc Thesis, City University, 1995. 96 pages.

Pushpakumara, D.D.A. Gooch. R.M. and Clarke. T.A. 1996. Real time image processing and data communications for a 3-D measurement system. 15th National Information Technology Conference Proceedings, The Computer Society of Sri Lanka, pp. 1-9.

Wang, X. & Clarke, T.A. 1996. An algorithm for real-time 3-D measurement. ISPRS Vol. 31. Part B5. pp. 587-592.

Wang, X. & Clarke, T.A. 1998. Separate adjustment of close range photogrammetric measurements. To be published in ISPRS International Symposium on Real-Time Imaging and Dynamic Analysis.

PAPER REFERENCE

Clarke, T.A., Li, L, & Wang, X. 1998. 3D-Net - a new real-time photogrammetric system. ISPRS Vol. XXXII, Part 5, pp. 29 - 34.