Real Time Tracking a Dynamic Object by Multiple Vision System

Amir Saeed HOMAINEJAD
Department of Surveying
The University of Technology of K.N.Toosi
No. 1346 Valli-asr Ave, Mirdamad Cross
Tehran 19697, I.R. IRAN

Commission V, Working Group IC WG V/III

KEY WORDS: Real Time, Processing, Expert System

ABSTRACT

This paper outlines a strategy that is designed for a multiple vision system consisting of four cameras, two of them with a convergent axes and a fixed position relative to the testfield and the other two cameras have a stereo position and normal relative to the testfield but their position are unknown. Basically, this test has been undertaken to detect and precisely track the movement of the object in the depth coordinates (in the Z direction) and compare with the outputs of a stereo vision system. The dynamic object is a doll that is activated by sound and has a vibration movement. The vibration movement is not predicted; therefore, the approximate trajectory of the movement cannot be introduced to the knowledge base as constraints for facilitating the processing. The results show the system is reliability able to detect and track the object. The processing time for detecting object from image is less than 30 msec, and the positioning precision is about 1\100000 on the image.

1. PREVIOUS TASKS

1.1 A Stereo Vision System for Tracking Dynamic Objects

A stable vision system consists of a stereo CCD camera was designed to acquire stereo images from a testfield. The vision system was set up about 2.2 m from the testfield. Images were acquired off-shelf and stored in the buffer. An algorithm was developed to retrieve sequence images from buffer in order of image acquiring. The algorithm was supplied with an expert system for detecting and extracting template target so that the vision system defined the position of itself and could compute the position of any object on the testfield (Homainejad and Shortis 1995a). Two different dynamic objects were separately moving around the testfield while some stereo images were taken. In the post processing, the algorithm detected objects based on the subtraction later images from the original image (Homainejad and Shortis 1995b). The first object had a simple shape and its movement was linear. The second object had complex shape and its movement was non-linear. For detecting and extracting the first object from images, the algorithm tested the subtracted images by a threshold. If value of pixels of an area on the subtracted images were satisfy the defined threshold, the algorithm detected that area as an object on the testfield. As an advantage of this strategy, the algorithm detected a thick area which satisfy the threshold so that noises could not be detected. For achieving this aspect, a few constraints about the object and its path of movement had been given to the algorithm. According to the defined strategy, the algorithm tracked the object.

For detecting and tracking the second object, another strategy was used because the path of movement of the object was non-linear and a specific point on the object should be detected. Therefore, a number of constraints about the path and the point which should be detected were given to the algorithm. In order to reduce the processing time, images were subtracted partially. Consequently, introduction of constraints to the algorithm was very important. The method of object detecting was the same of the method of the first method. In this method, the algorithm detected a thick area. Then, the algorithm detected the point on the image according to constraints and a knowledge base. The algorithm according to constraints chose an area which should be investigated and detected the point.

As mentioned earlier, a specific expert system was developed for achieving the aspects of these tasks. Because a few codes were used in the expert system, the expert system was developed in C language. Aims of developing the expert system were:

- to recognise and extract template targets,
- to define the position of control points on the testfield,
- to recognise and detect the object in the testfield, and
- to track a specific point on the object.

Offcourse, a knowledge base was introduced to the algorithm for achieving above aspects. The knowledge base was about the shape and the colour of the object, the colour of the background, template targets, and function of the path of movement of the object.

Remarkable results were obtained from these tasks. Detecting, extracting, and tracking the first object was done successfully in real time processing. the specific point could detected from the second object. The absolute accuracy was about sub-pixel.

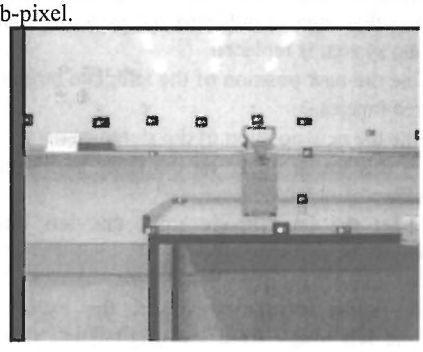

Figure 1: Illustration of the first object.

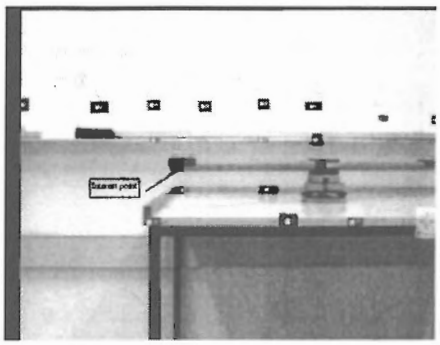

Figure 2: Demonstration of the second object and interest point.

1.2 Real Time Tracking Objects by Unstable stereo vision system

The previous section explained two tasks of real time processing of tracking objects. This section will explain another task which is sought answer to some basic questions.

- At what processing time can the system track a dynamic object?
- Is the program able to reposition the vision system, when the system is relocated?
- To what extent is the method robust and reliable?

To address the questions, the vision system was setup in the front of the testfield at the distance of about 3 meters. Then, the system acquired a stereo image from the testfield. This stereo image was permanently stored in the buffer as original images. The original images were regularly compared with the later images for controlling the displacement the vision system. The vision system was relocated in a new position along the Z axis. Following this, the vision system acquired stereo images from the testfield, while an object was moved throughout the scene. Only one object was used in this test, but two different tests were carried out. It should be noted that the colour and the shape of the object made the processing complicated.

A dialogue based on the strategy was designed to be introduced into the expert system. The dialogue's aspects were:

- subtract the left image of the later image from the original image,
- analyse the area of a template target on the image, and if that area is not zero the system assumes the vision system is replaced,
- define the new position of the template targets on the stereo images,
- define the new position of the stereo camera,
- store this stereo image on the array permanently for tracking,
- subtract the later image from this left image for tracking the dynamic object.

In order to implement these aspects, the expert system subtracted the left image from the original left image, and investigates the area of the template target in the subtracted area. Figures 3, 4, 5 present the original, the later and the subtracted images, respectively. As Figure 5 presents, the template targets on the subtracted image are not zero, and can clearly be identified from the background.

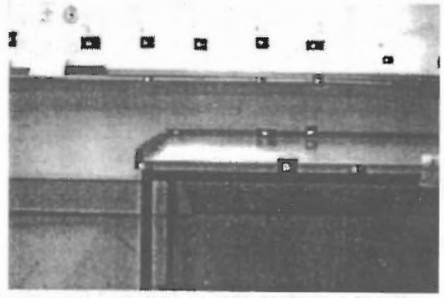

Figure 3: Illustration of the first left image.

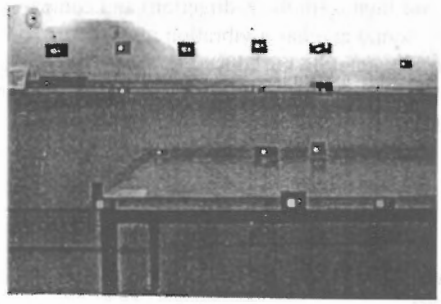

Figure 4: Illustration of the left image when the stereo vision system was in the new position.

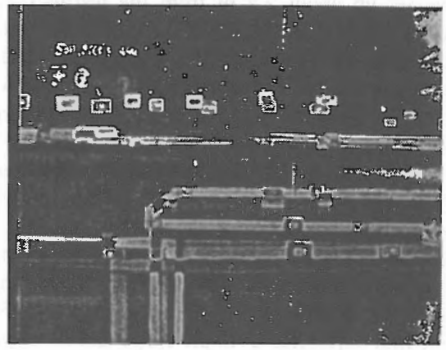

Figure 5: Illustration of the subtracted image.

With analysing Figure 5, we can understand bright patterns show testfield from two different position. This result was the main skeleton of the strategy of vision system positioning by the expert system. Based on this result, the expert system investigated a number of patterns, which were bright, on the subtracted image. Basically, the expert system was looking for template targets so that it could define the new position of the vision system.

After positioning the vision system, the expert system continued to search and detect a dynamic object in the testfield. The method of searching and detecting is already explained in previous Section. This means that the expert system using the partial subtraction defined the position of the object. It should be noted that the method of subtraction was modified for this test because the expert system could successfully track the object. In addition to subtracting, the system in this test compared corresponding pixel values of

the images because the colour of the object was white and was very close to the background colour; otherwise, the subtraction would not be successful (Homainejad 1996). Consequently, the comparing of the two images was necessary. For comparing, the program compared the pixel values of two images, then subtracted the image with the small pixel values from the image with the big pixel values. According to this method, the expert system compared the common area from the two images and, when it recognised a significant difference between corresponding pixel values of two images, it would confirm that an object was location area. Then, the expert system defined the position of the centre of the gravity of that area from the stereo image. Finally, the system defines the position of that point in the object system.

The object in this test had a complicated shape. The object was a bottle with a waist. The depth of the waist was not more than one centimetre. Therefore, it was decided to track two points on the object; one point in the waist area and other point in else where. Then the outputs were investigated and analysed. Hence, the expert system tracked the object twice. The second test was fulfilled to confirm that the expert system was able to track a correct point upon the object.

Because the colour of the object was bright and white, the detection of the correct area was very difficult. Therefore, a common area on the images was selected in order that the expert system could regularly control the lighting for each image. If a difference was found the expert system defined a scale factor to apply in the images. Equation (1) explained the method of defining the scale factor. It should be noted that the proposal of defining the scale factor was based on the assumption of a unique light illuminating the object during of each period of image acquiring.

$$s = \frac{p_m^l}{p_m} \qquad (1)$$

$$P_i^l = s \times p_i^l$$

where:

p_m^l is the mean value of the pixels of the common area in the later image,

p_m is the mean value of the pixels of the common area in the original image,

p_i^l is the pixel value in the later image,

P_i^l is the pixel value in the later image after applying the scale factor,

s is the scale factor.

The histograms of common area on two images were investigated so that Equation 1 could be successfully applied on images.

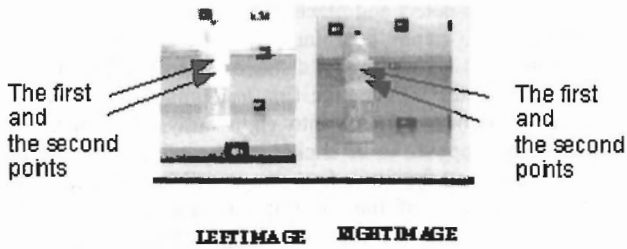

Figure 6: The first and the second point upon the object

2. Tracking a Dynamic Object by Multiple Cameras Vision System

2.1 Proposed Tracking

This section addresses the proposed method of tracking for this project. The structure of tracking is based on the aspects below:

- real time processing,
- automatic processing,
- reliable output,
- precise output,
- the coordinates are computed in a real world coordinate system,
- the system is able to position/reposition itself.

A method was developed in this project was not to be influenced by issues. These issues include unknown position of the camera (cl Negahdaripour et al 1991, Sundareswaran 1991), the effect of noise in depth (cl Oliensis and Thomas 1991). In addition, the method needed to achieve the necessary aspects of tracking. The method had to be able to automatically recognise a dynamic object in the scene and track it in real time. The output needed to be reliable, and precise. In addition, the method needed to be able to position the vision system when the system was relocated. Therefore, the single camera system method was strongly rejected for this project because of its lack of performance. The lack of performance is a result of the inaccurate computation of depth, a limitation that almost all the single camera system methods suffer from. Homainejad (1997) addressed all these issues. As a result, a multi camera system method is selected in order to achieve the aims of this project. The adopted system consist of a multiple cameras configuration, a frame grabber, and a computational program that was developed specifically for this project. A multiple cameras approach is less sensitive to noise than a single camera method, because random noise in a sequence images is not similar, and the precision of the output is increased according to the inverse ratio of $\sqrt{n}$, where, n is the number of images.

2.2 Strategy of Tracking a Dynamic Object by a Multiple Vision system

This section presents a strategy that is designed for a multiple vision system, consisting of four cameras, two of them with a convergent axes and a fixed position relative to the testfield and the other two cameras have a stereo position and normal relative to the testfield but their positions are unknown. Figure 7 shows the chart of the multiple vision system and the test field. The strategy of tracking in this system is quite different from those which are already explained in the previous two sections. Basically, this test has been

undertaken to detect and precisely track the movement of the object in the depth coordinate (in the Z direction), and compare with the outputs of a stereo vision system. In order to achieve these aspects, the first four images of the vision system will be registered into eight arrays, of which four arrays are supposed to be unchanged during the entire of the processing, and the other four arrays will be changed. After the computation of the exterior parameters of the stereo camera by using the method of Homainejad and Shortis (1995a), the expert system will detect and track a dynamic object from a sequence of images that have been captured by the multiple vision system, and stored in the buffer.

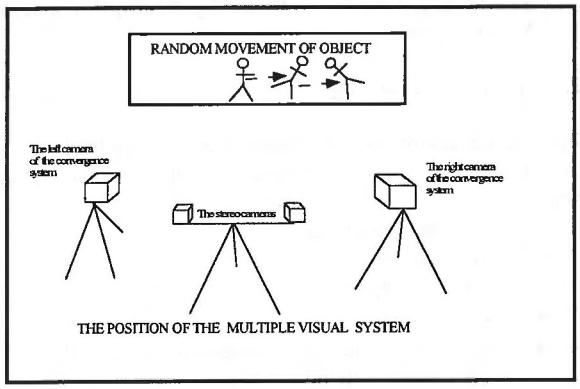

Figure 7: The diagram of the multiple vision system.

The dynamic object is a doll that is activated by sound and has a vibration movement. The vibration movement is not predicted; therefore, the approximate trajectory of the movement cannot be introduced to the knowledge base as constraints for facilitating the processing. The expert system will read and register a part of the left image of the stereo pair into one of the temporary arrays. The array already has stored the first left image. The expert system replaces the new pixels with old corresponding pixels. Therefore, this array includes two images of which that the second image partially is replaced with the first image.

Unlike the strategy of object detection that was explained in the previous two sections, the expert system does not subtract the first left image of the original stereo pair from the left image of the later stereo images, for finding an object in the next images. In contrast, the expert system compares the first and the later images of the stereo image for avoiding the interference of the background into the subtracted images which is a common problem for all configurations. The comparison process is that the expert system compares each pixels in the later image with its corresponding pixel in the original image, if there is not different the old pixel will be left there, otherwise the old pixel will be replaced by the new pixel.

It should be noted that the expert system will implement another process, before tracking the dynamic object. The expert system will define the parameters of the bilinear transformation of the left image of the stereo image and the other three images. This procedure will be done to transfer detected points from the left image of the stereo image to the other images quickly and precisely. Therefore, any point that is extracted from the left image of the stereo image can be mapped on to the other images.

When the expert system detects a point upon the object from the left image of the stereo image, it maps that point on the other images according to the parameters of the transformation and extracts the position of the point from those images. This processing guarantees the detection and extraction of the precise position of a common point on four images. In addition, it determines a small area of the other images that are needed in order to be registered into the other changeable arrays. Then, the expert system will register those areas into the other arrays and extract the position of the common point. Finally, the expert system will calculate the position of the point in the object coordinates system and track the object. Figure 8 demonstrates the chart of the strategy of the object detecting for multiple vision system.

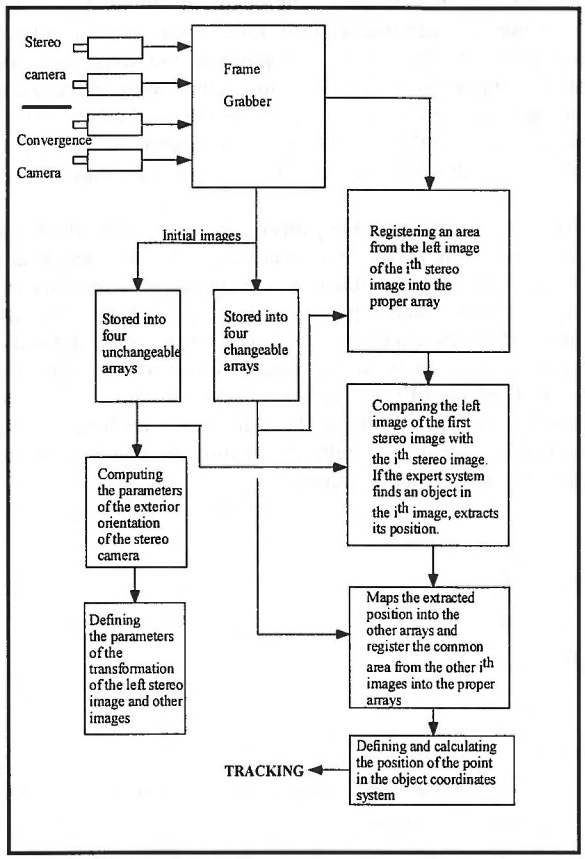

Figure 8: The chart of strategy of tracking a dynamic object by the multiple vision system.

2.3 Evaluation of Proposed Strategy

This section presents the results of tracking a dynamic object by a multi cameras vision system. The strategy of tracking for this test is different from the two previous strategies. It is assumed that the image registration on an array belongs to the initialisation of the expert system; additionally, this assumption provides at least two advantages. The first advantage is that the whole of the original images can be registered on arrays, and consequently any part of each image can be quickly analysed. In addition, the second advantage is that the expert system saves a significant processing time for image processing because reading and writing data from the array is much faster than from the file.

Another difference between this strategy and the two previous strategies is related to the object detection method. As mentioned before, the process of the object detection of two previous methods was based on the detection of a point upon the object on the left image and the definition of its correspondence point on the right image by using the epipolar method. Those methods were designed based on the fact that the stereo vision system had a normal position relative to the testfield, and the cameras' principal points were located along a line parallel to the X axis. Therefore, the correspondence of a point on the left image can be easily defined on the right image by using the epipolar method. However, this method can not be used for this test because the four cameras are not positioned along X axis. In other words, the cameras' principal points are not located along a line. Consequently, those methods cannot be used for this test.

The strategy of object detection for this test is based on detecting an interest point on the left image of the stereo image and transferring that point to the other image system of the vision system for defining its corresponding points on the other images. The bilinear method is used for object transformation.

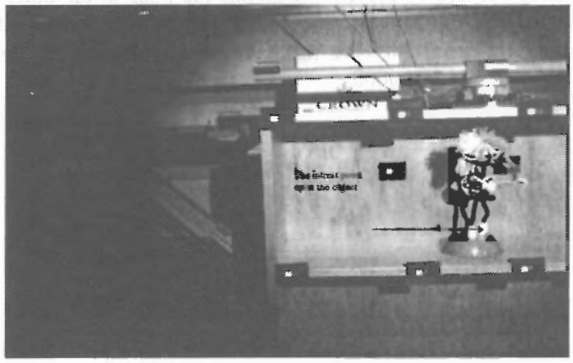

Figure 9: The demonstration of the third object.

Another difference of the object tracking strategy of this test with two previous tests is that the method of the subtraction is replaced by the method of comparison. There are at least two reasons for changing this method. The first reason is that the object does not have a unique colour, and the second reason is that the movement of the object is not predictable. Figure 9 demonstrates the object that was used for this test, a doll that is shocked by sound; and hence the next position of each part of the body cannot be predicted.

The method used compares the pixel value of an area of the original image with its common area in the later image. If a difference is detected, the expert system will compare that area with a knowledge base that is already stored in the buffer. The knowledge base defines information about the object and the interest point upon the object. The interest point is a point in the right foot of the doll as Figure 9 demonstrates. The equation for comparison method is explained below.

$$\left|p_i^l - p_i^o\right| \rangle m, m \neq 0$$
$$a \langle p_i^l \langle b, a \langle b \neq 0 \qquad (2)$$

where:

p_i^l is the pixel value of the later image,

p_i^o is the pixel value of the original image,

m is the threshold value,

a, b are the minimum and the maximum values of the second threshold.

This method enables the system to detect the interest point on the object even if the point is located in a difficult area. In addition, this method overcomes the subtraction problems, such as non-zero area and interference of background to foreground, that the two previous tests were suffered from.

In order to achieve the aims of this test, a dialogue was designed and introduced to the expert system. The dialogue's orders are:

- detecting the interest point by comparing the left original image of the stereo image and the left image of the later stereo image,
- transferring the detected point to the other images,
- detecting the corresponding points on the other images, and
- defining the position of the point in the object coordinates system.

Subsequently, the expert system follows the orders. It is necessary to mention that the parameters of the bilinear transformation from the left image of the stereo image to the other images are calculated automatically in real time. When the stereo matching is fulfilled, the expert system calculates these parameters because the coordinates of four common points of the four images are already determined. The processing time of computation is less than 1 ms. Therefore, each point on the left image of the stereo image can be transferred to the space of the other images according to these parameters.

It should be noted that epipolar method is useful for vertical imagery. The above parameters can be obtained if their corresponding and scale factors are known on the object. Figure 10 demonstrates this situation. Because the object was vibrated and its foot had a new position regularly, the above equation could not be used in this test. Figures from 11 to 14 show a sequence of the object images. As this figures demonstrate, the position of the foot of the object is not along a regular direction. Consequently, it was decided that after transferring the point to the other image spaces, the expert system open a search window on the other images and detect that common point. Finally, the expert system would define the position of the point on the object coordinate system.

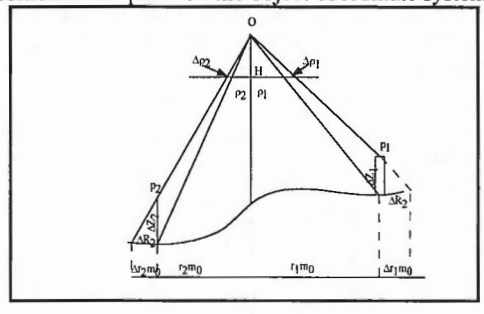

Figure 10: The demonstration of the distortion of the image along the depth of the field.

Figure 11: The first left stereo image of the third object.

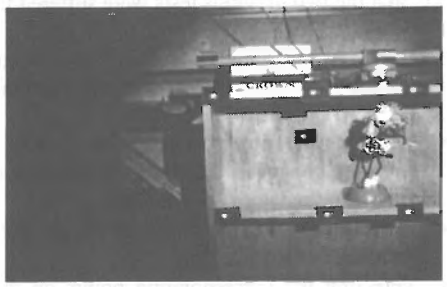

Figure 12: The second left stereo image of the third object.

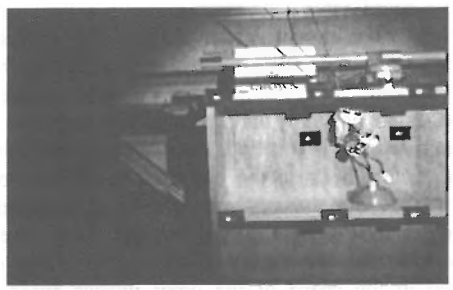

Figure 13: The third left stereo image of the third object.

Figure 14: The fourth left stereo image of the third object.

The coordinates of the foot of the object were observed with two total stations using the intersection method, and their RMS error are less than 0.01 mm. The coordinates are listed in Table 1. This table shows the coordinates of a point on the right foot of the doll for different positions.

Position	X	Y	Z
1	5.5332	4.3961	-4.0573
2	5.5302	4.3951	-4.0662
3	5.5604	4.4147	-4.055
4	5.5445	4.4251	-4.0646

Table 1: The coordinates of the interest point on the object. Units are meters.

The computed coordinates of the point of the object are listed in Table 2. The expert system defines the coordinate of a point on the foot between the toe and the angle.

Position	X	Y	Z
1	5.5298	4.3936	-4.0521
2	5.5299	4.3902	-4.06
3	5.56	4.4103	-4.05
4	5.5499	4.4202	-4.06

Table 2: The obtained coordinates of a point on the foot of the object by the algorithm. Units are meters.

The above coordinates were obtained by using the intersection of the rays. Comparing the two tables, the ratio of two coordinates in the depth is about $\frac{1}{1000}$ for each single observation, which is a good result according to the depth field of the camera and method of object detection. Figures 15, and 16 present the graphs the trace of the object in three plane of XY, and XZ planes, respectively. Finally, it should be noted that the precision of coordinates on the image is about 1:100000.

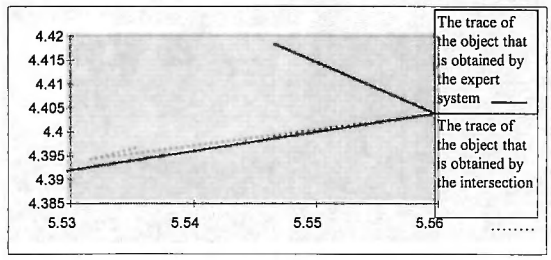

Figure 15: the graph of the traces of the trajectory of the object in XY plane. The units are meters. The graph compares the obtained results by the surveying intersection and the expert system. As the two traces demonstrate, the tracking of the expert system is parallel with the survey intersection and only the last point is exactly coincidence with the observed point. units are meter.

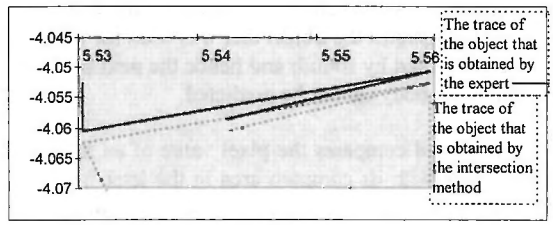

Figure 16: the traces of the two tracking in the XZ plane. The units are meter. The top trace relates to the tracking by the expert system method and the second trace relates to the surveying observation.

In addition, the processing times of the processing are presented in Table 3.

The time delay for registration of four 768x484 images on eight arrays	3.25 Sec
The time delay of detecting and extracting of four template targets from the stereo image, and stereo matching	10 ms
The time delay of detecting, extracting, tracking the object from four images. The time delay relates to the positioning of the four template targets upon the object	10 ms

Table 3: The processing times of the three different processes of the multiple vision system.

These results verify that the most expensive processing belongs to the image registration on the arrays. Therefore, two suggestions can be given for improving the processing time. The first is to improve and develop the image library for photogrammetry purposes and, in particular, for the real time processing. The second suggestion is to register the images on an array instead of a file during the image acquisition. The second suggestion can be fulfilled when the vision system is concurrently connected to the computer and the object is tracked on-line.

In addition to the above test, the object was tracked by the stereo vision system and its results are presented in Figure 17. This additional test was carried out to compare the differences of tracking between a stereo camera vision system and a multiple cameras vision system. The outputs of the stereo cameras vision system are very close to the other system, and the stereo camera vision system can be used for tracking the object when less accuracy is required. In addition, the second method can be used for certain applications of the object tracking when the object is moving along a defined direction.

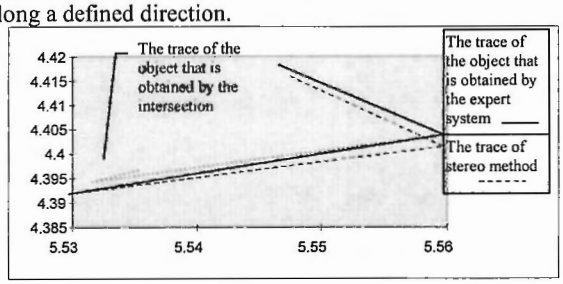

Figure 17: The graph of three methods of the tracking in XZ plane. The units are meters. The trace of the tracking by the stereo vision system is close to the multiple vision system. Therefore, the stereo vision system can be used for the applications that do not need to have a high precision.

The processing times of the stereo vision system is presented in Table 4.

The time delay for registration of two 768x484 images on four arrays	1.64 Sec
The time delay of detecting and extracting of four template targets from the stereo image, and stereo matching	10 ms
The time delay of detecting, extracting, tracking the object from the stereo images. The time delay relates to the positioning of the four template targets upon the object.	10 ms

Table 4: The Processing times of the three different processing for the stereo vision system.

The above results verify that the processing time for object detection and tracking is not different from the multiple camera method. The processing time of image registration is significantly slow for the multiple method. If the processing time of image registration is excluded from the whole of the processing time, then the processing can be fulfilled in real time without sacrificing precision.

3. SUMMARY

This paper presented new method of dynamic object tracking, which was fulfilled in real time. The processing time of each processing was less than 30 msec. Additionally, the decision to use a multiple camera vision system was aimed at acquiring precision. Two different vision systems were used in this research; a vision system which included stereo CCD cameras, and had a normal position relative to the testfield; and one which included four CCD cameras. Two of them had a stereo position, and the other two cameras had convergence positions relative to the testfield. The depth of the testfield was about 30 cm.

Certain problems distorted the output images. Following problems were encountered during the acquisition of the imagery:

- lighting,
- target texture,
- background colour, and
- object colour.

The lighting was the most important problem because it degraded the output images. In addition, the background colour was white and bright that interfered the foreground and the template targets. Additionally, the texture of the template targets was very bad because the black patterns reflected the light that interfered the control points and symbols. In order to overcome these problems, firstly it was decided to use a unique light for all tests. Secondly, a common area was selected in the sequence images for defining a scale factor for overcoming the lighting problem.

The use of template targets has many advantages. For example, they can be used for automatic stereo matching. In addition, these template targets can be used for automatic orthoimagery and image transformation for which an example was presented in Section 2.3.

In addition to the above advantages, the template targets can be used for automatic vision system positioning. For vision system positioning, sequence images are subtracted from the original image, and, if double template targets exist in the image, the expert system can recognise the displacement of the vision system. This strategy was used for re-positioning the vision system in the second test of the stereo camera vision system.

Additionally, in order to achieve real time tracking of a dynamic object, an expert system was developed. The expert system partially subtracted sequences of images from the original images. Then, it detected the object from the subtracted image. When it found a non-zero pattern in the subtracted image, the expert system recognised it as the object. The expert system then positioned a point upon the object. In the first test of tracking, two different objects were

used. The expert system determined a point upon the first object which was located in the centre of the gravity of that part of the object in the subtracted image. In addition, the expert system searched for a particular point in the second object. In order to detect this point, a certain knowledge base describing the object and the interest point was introduced to the expert system. As a result, the expert system could successfully recognise and extract the interest points upon two objects. It should be noted that the interest point on the left image of the stereo image was detected according to the above explanation, and for detecting the corresponding point on the right image the expert system used the epipolar line method.

The method of interest point detection was developed for the second test of tracking. In this test, the expert system simultaneously subtracted and compared the images because the image colour was very white and bright and very close to the background colour. In order to achieve a good object recognition from the background and overcome the lighting situation, a common area was selected in the images for defining a scale factor for lighting. This scale factor was applied on the images. The object in this test had a complicated shape and the expert system detected two different points. One point was in the waist area and the second point was in the other area. The expert system could successfully recognise and detect these points on the left image of the stereo image and on the right image using the epipolar method. Next it tracked the points. In this test, the images were partially subtracted and compared as well, to save the processing times.

The third test of tracking used an object that was activated and re-positioned by sound. Therefore, the expert system had no knowledge about the next position of the object and could not predict its new position. Consequently, the processing was more complicated than for the two previous tests. In this test, the object was tracked by both the multiple cameras vision system and stereo camera vision system. In addition, two points upon the object were tracked by the survey intersection method. The expert system calculates the parameters of the bilinear transformation between the left image of the stereo pair and the other images, in order to define the position of interest points in the other images very quickly. The results of tracking with the expert system and the survey intersection were very close to each other, and it confirmed that the method can successfully recognise, detect and track a point upon a dynamic object.

4. REFERENCES

Homainejad, A. S. (1997). Real Time Photogrammetric Processing. Department of Geomatics, The University of Melbourne, Australia

Homainejad, A. S. (1996). Real Time Tracking of a Dynamic Object. XVIII CONGRESS of International Society for Photogrammetry and Remote Sensing, Vienna, Austria, ISPRS, 243-246.

Homainejad, A. S. and M. R. Shortis (1995a). Development of a Template for Automatic Stereo Matching. ISPRS Intercommision Workshop: From Pixels to Sequences, Zurich, Switzerland, 318-322.

Homainejad, A. S. and M. R. Shortis (1995b). Stereo Vision System for Tracking a Dynamic Object. SPIE, Videometrics IV, Philadelphia, Pennsylvania, SPIE- The Internation Society for Optical Engineering, 264-271.

Negahdaripour, S. and S. Lee (1991). Motion Recovery from Image Sequences using First-Order Opticaal Flow Information. IEEE Visual Motion, Nassau Inn, Princeton, New Jersey, 132-139.

Oliensis, J. and J. I. Thomas (1991). Incorporating Motion Error in Multi_Frame Structure from Motion. IEEE Visual Motion, Nassau Inn, Princeton, New Jersey.

Sundareswaran, V. (1991). Egomotion from Global Flow Field Data. IEEE Visual Motion, Nassau Inn, Princeton, New Jersey, 140-145.