TOWARDS AN OPERATIONAL DIGITAL VIDEO PHOTOGRAMMETRIC SYSTEM FOR 3-D MEASUREMENTS

E. Tournas, Postgraduate Student
Dr. A. Georgopoulos, Assistant Professor
Laboratory of Photogrammetry
National Technical University of Athens
Greece

Commision II, Intercommission Working Group II/III

KEY WORDS: Digital Video, 3-D Measurements, Hardware, Software, Industrial Application

ABSTRACT

In this paper the first attempt for developing a complete and operational photogrammetric system for determining the trajectories of moving objects in space using digital video is described. The system is based on a Personal Computer and consists of off-the-shelf hardware in order to enable the stereoscopic observation. The input data is a sequence of synchronised image video frames acquired with two video cameras. It is usually assumed that the changes in a scene are due to object motion and the two cameras are stable. The system extracts masks of moving targets in order to recognise them, thus determining 3-D coordinates of the observed points and subsequently producing the motion characteristics of the object.
The integration of the system involves development of specialised software. Firstly camera calibration is performed on both cameras with the help of a self calibrating least squares bundle adjustment solving simultaneously for all the unknown parameters. Secondly the relative and absolute orientations are performed with the help of a two camera triangulation algorithm, while the image coordinates of the observed points are determined automatically by using a cross correlation algorithm which matches the observed premarked targets with a predefined target template. Stereoscopic vision is ensured with the help of the CrystalEyes system applied on resampled data, which form a pair of epipolar images. Finally specialised software has been developed in order to animate and view stereoscopically image sequences, perform 3-D measurements and support interactive point and line drawing. In order to thoroughly investigate the suitability of the methodology an application involving analysis of seismic movements of scaled models was performed. Experiments have been carried out in the Laboratory of Earthquake Engineering of NTUA, in order to monitor the movements in space of an ancient Parthenon pillar model of 1:3 scale. The results of this practical application are also briefly presented and discussed.

1. INTRODUCTION

The determination of 3-D positions of moving objects is a highly important subject in many industrial processes and research areas. Vision based systems combining the principles of machine vision and digital photogrammetry have great potential for solving such kind of measurement problems. In the following the first approach for developing a digital stereoscopic video system for determining the trajectories of moving objects in space will be described.

The main aim of the system developed was to provide a way of monitoring, recording and determining the, usually unpredictable, motion of objects under seismic action. The results should satisfy certain tight accuracy limits, while the cost should, as always, be kept to a minimum. At the same time simplicity and applicability was of high significance.

Since rapid and unpredictable movements ought to be recorded, the use of video as the only suitable data capture means was inevitable. When digitised, this would produce a huge amount of electronic data, whose storage and processing would be very difficult. Hence the need arose for developing a custom made system in order to be able to metrically exploit the available data. The development of the system called for specialised software in order to exploit the off-the-shelf hardware which was used. The system in its present stage of development consists of a 486 PC with a 17" monitor capable of producing refresh rates at a frequency of 120 Hz and the necessary hardware for the stereoscopic observation (Figure 1).

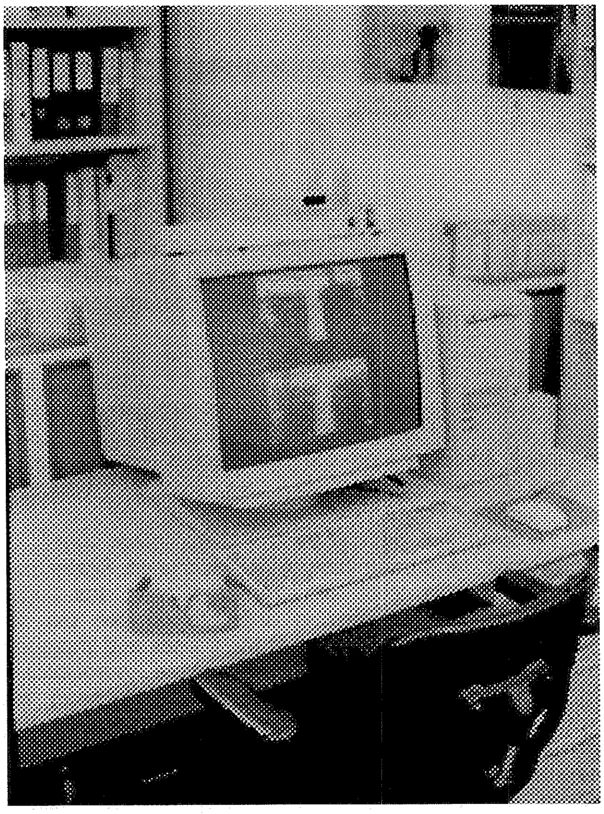

Figure 1

2. SOFTWARE AND HARDWARE ASPECTS

For the integration and operation of the system specialised software was developed, involving all photogrammetric problems, for the adequate metric exploitation of the video data.

2.1 Camera calibration

Camera calibration of the video camcorders used is obviously of utmost importance, in order to ensure high accuracy for the measurements. It may accurately be performed with the use of a self-calibrating least squares bundle adjustment. Each camera setup introduces 15 unknowns into the adjustment. Six of them are station dependent unknowns, and are usually referred to as exterior orientation parameters. These are the three space coordinates of the perspective centre (X_o, Y_o, Z_o) of the camera lens and the three rotation angles (ω, φ, κ) of the lens axis. Since the commercial camcorders used could not be considered as metric, each camera introduces 9 additional station dependent unknowns. The set of the additional unknowns includes the location of principal point (x_p, y_p) on the image plane, the camera constant (c), two parameters for radial distortion, two for decentering distortion and two affine transformation parameters. The mathematical model is based on the well known photogrammetric collinearity equations enhanced with additional correction terms Δx, Δy, in order to compensate for camera lens distortion errors (He *et al.* 1992, Snow *et al.* 1992):

$$\begin{aligned} \Delta x &= x_p + \bar{x}(r^2-1)a_1 + \bar{x}(r^4-1)a_2 + (r^2-2\bar{x}^2)a_3 + 2\bar{x}\bar{y}a_4 + \bar{x}a_5 + \bar{y}a_6 \\ \Delta y &= y_p + \bar{y}(r^2-1)a_1 + \bar{y}(r^4-1)a_2 + 2\bar{x}\bar{y}a_3 + (r^2-2\bar{y}^2)a_4 - \bar{y}a_5 \end{aligned} \quad (1)$$

where:

x_p, y_p	the image coordinates of the principal point
x, y	the image coordinates with respect to the principal point
r	the radial distance of the point from the principal point
a_1, a_2	radial distortion parameters
a_3, a_4	decentering distortion parameters
a_5, a_6	affine transformation parameters

In order to enable the algorithm to solve for the unknown parameters, a test field including at least 8 suitably signalised control points is necessary. These control points should be stable and placed surrounding the moving object. Their location in space is determined with the help of precise geodetic measurements.

For the determination of the control points' image coordinates, an automatic target detection algorithm is used. Firstly image processing tools, such as average filtering and contrast stretching, may be used in order to remove the noise introduced during the digitization of the frames and improve the quality of the images. Initial target image coordinates are determined by using a cross correlation algorithm which matches the observed targets with a predefined target template (Vernon 1991). It has been found thet the use of circular black targets on white background facilitates this phase of the work, because they are not distorted considerably from different angles of view. To improve the accuracy, a centroid detection algorithm has been finally used, which computes the centre of gravity of the black circles with sub-pixel accuracy.

Object coordinates in 3-D space are determined by a two camera triangulation algorithm (Snow et al. 1992). Image coordinate measurements of signalised points of interest are performed automatically on synchronised stereoscopic video frames of the object, in order to avoid manual identification of the targets separately on each frame. Synchronisation is achieved by matching the frames with the same clock indication, which is present on each recordered frame. Absolute synchronization of the two cameras requires specialized hardware which is not usually available for commercial video camcorders.

The image coordinates of the observed targets are introduced into the computation as tie points with unknown positions. For each tie point 3 additional unknowns and 4 observation equations are added.

2.2 Rectification to the normal case

The production of epipolar images definitely simplifies the process of stereoviewing and measuring. The rectification of a stereo pair to the normal case is achieved by using epipolar geometry and image processing techniques. The basic idea is simple: if the left and right image planes are coplanar and the horizontal image coordinates axes are collinear (no κ rotation about the optical axes), then the epipolar lines are parallel in both images and the corresponding lines have the same y coordinate. Since such a condition is quite difficult to achieve in practice, a geometric transformation should be employed in order to transform both left and right images to the desired coordinate system of the common plane.

In the following, the algorithm used to rectify a stereo pair of video images to the normal case will be described. First of all, the elements of the relative orientation should be determined. Assuming that the left photo is fixed in position and orientation, i.e. the three translation displacements and three rotations are equal to zero, five relative orientation parameters have to be calculated. The translation displacement dX in the X direction is fixed at a value approximately equal to the image base, in order to establish the scale of the stereo model approximately equal to the image scale. The five unknown parameters are then computed for the right image. They are the three rotation angles dω, dφ, dk and the two translation displacements dY and dZ.

In the ideal normal case, the right image is considered as "vertical" (dω=dφ=dκ=0) and no displacements exist in Y and Z directions, i.e. dY=dZ=0. This means that the right image is parallel to the left one and the epipolar lines are parallel to the image coordinate fiducial axis system. Although this ideal case does not actually exist, it is possible to rotate a gemerally tilted right photo to a coordinate system that is parallel to the fixed coordinate system of the left photo by using the following relations (Chen & Scarpace 1990):

$$\begin{aligned} x'_r &= m_{11}x_r + m_{12}y_r - m_{13}c \\ y'_r &= m_{21}x_r + m_{22}y_r - m_{23}c \\ z'_r &= m_{31}x_r + m_{32}y_r - m_{33}c \end{aligned} \quad (2)$$

where:

x_r, y_r	right image coordinates in coordinate system of the tilted image
m_{ij}	elements of rotation matrix

Since this transformation produces a different z'_r for each image point, the right image must also be transformed to the common plane. In this study, a common plane at (-c-dZ) has been used and each pair of rotated image coordinates x'_r, y'_r is transformed to this common plane, using the following equations (Chen & Scarpace 1990):

$$Vx'_r = (-c-dZ)\frac{x'_r}{z'_r}$$
$$Vy'_r = (-c-dZ)\frac{y'_r}{z'_r} \qquad (3)$$

Both images are now free from tilt, but because of the Y displacement of the right image, the epipolar lines are not parallel to the row direction of the scanning coordinate system. Since the dY translation parameter of the right image is known, a new coordinate system is generated, where the effect due to Y translation is eliminated,. This is done by rotating the left fixed coordinate system by an angle θ, equal to the angle between the X axis and the line connecting the principal points of the two images. The angle θ is given by the following equation:

$$\theta = \tan^{-1}\left|\frac{dY}{dX}\right| \qquad (4)$$

The coordinates on the left image in this coordinate system become:

$$Fx_l = x_l\cos\theta + y_l\sin\theta$$
$$Fy_l = -x_l\sin\theta + y_l\cos\theta \qquad (5)$$

Similarly the coordinates on the right image become:

$$Fx_r = Vx'_r\cos\theta + Vy'_r\sin\theta$$
$$Fy_r = -Vx'_r\sin\theta + Vy'_r\cos\theta \qquad (6)$$

All the above equations describe a direct transformation, which performs the production of epipolar images, i.e. the rectification of a stereo pair to the normal case. In practice an inverse transformation is actually used according to the following methodology (Jain et al. 1995). First of all, the locations of the four corners of each image on the common plane are determined. Then, new left and right image grids are created with grid cell dimension equal to the scanned pixel size. Finally, each grid point of the new image is transformed back to the original image. Bilinear interpolation is used to interpolate pixel values to determine the grey tone values for the new left and right images in the common plane. Before interpolation, the computed image coordinates must be "corrected" by reintroducing radial lens distortion and other systematic errors known from camera calibration. This is necessary because the calculation of the rectified image coordinates considers only "pure" undistorted values, while the actual location of a point on the image is affected by the systematic errors introduced by the interior orientation of the video camera.

2.3 3-D Animation of Stereoscopic Images

When the rectification to the normal case is completed for all captured image frames, a 3-D virtual representation of the recordered event may be realized on the computer screen. The key to display stereoscopic images on a single flat screen is rapid alternation between left and right images, while ensuring, at the same time, that each image reaches only the intended corresponding eye. Specialised hardware available from Sterographics Corp. was used in this study in order to display 3-D image sequences on the computer screen.

The system, called CrystalEyes, consists of an eyewear with shuttering lenses and an infrared emitter. The emitter, which sits on the top of the monitor, is connected to the display hardware and broadcasts a synchronization signal to the eyewear. The eyewear receives the signal and rapidly directs the appropriate image to the corresponding eye. When the left image is on the video screen, the left lens opens while the right lens closes. As a result, the viewer perceives a true stereoscopic view. Research has shown that for a stereoscopic view without any annoying flicker, the computer monitor and shutter lenses should be able to alternate the display of each image 60 times per second. This, however, is not the case in most SuperVGA boards, which do not support refresh rates greater than 60-72 Hz in high resolution graphics modes. This means that, at best, the monitor and shutter system will be able to deliver 30 images per second. To solve this problem, a separate GDC3 video converter is provided by Sterographics which takes the 60 Hz signal from the video board and converts it to 120 Hz.

Although the CrystalEyes system is simple and does not require any hardware modification to the computer, specialised software has to be developed in order to map stereoscopic views on the screen in the desired format. According to the CrystalEyes standards, the left eye views must be placed to the top half of the screen and the right eye views to the bottom half. The two images, called subfields, are vertically compressed by a factor of two (Lipton 1991). This is necessary, because when the system passes in stereo mode, both left and right subfields are vertically interlaced and alternatively displayed on the screen. Another important point to add, is that when the two images are interlaced, there are some horizontal lines which become invisible in both images. These lines are between 10-20 and are placed at the end of the left and the top of the right image. To overcome this problem, a blank horizontal area must be inserted between the two subfields (Figure 2). The height of this area depends on the monitor type and on the graphics resolution. A calibration procedure is necessary in order to determine the height of the blank interval before any image is mapped to the screen in use.

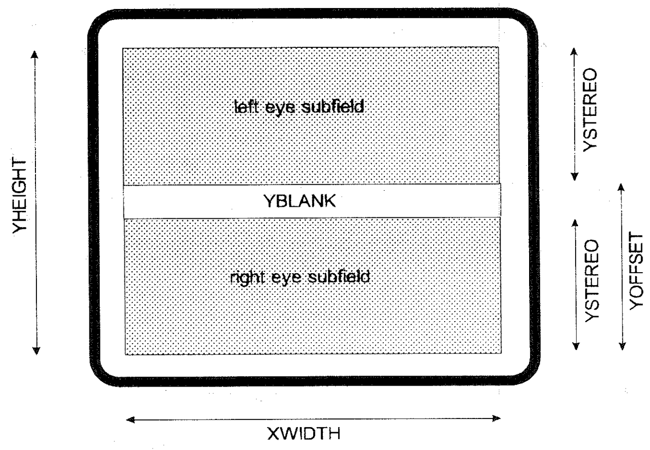

Figure 2

The representation of a dynamic scene in 3-D mode deepens on the display of a series of images, which consists of two parallel perspective projections. According to the CrystalEyes standard, the screen is divided in two rectangular viewports.

Since the video frames are usually captured in a 728x568 pixels resolution, a display resolution of 800x600 pixels is good enough in order to display vertically the two subfields interlaced, plus the required blank interval between them. The images are loaded from the hard disk, placed to the system RAM memory and then moved on to the display memory.

From 22 to 30 stereoscopic views per second should be displayed on screen in order to achieve smooth animation. This is rather impossible for most PC based systems, because of the large amount of digital data required. About 110 MB of image data should be loaded from disk, vertically compressed, interlaced and moved to the display memory in order to be able to produce 5 seconds of 3-D animation at this frame rate. At present all the above described operations allow for about 5-6 views per second, which results to a slow-motion 3-D animation (Figure 3). A significant improvement can be done by suitably programming the multimedia graphics accelerators which are present in some video boards. In this way, image scaling and bit block transfers between system RAM and display memory will be undertaken by the accelerator, independently of the system processor.

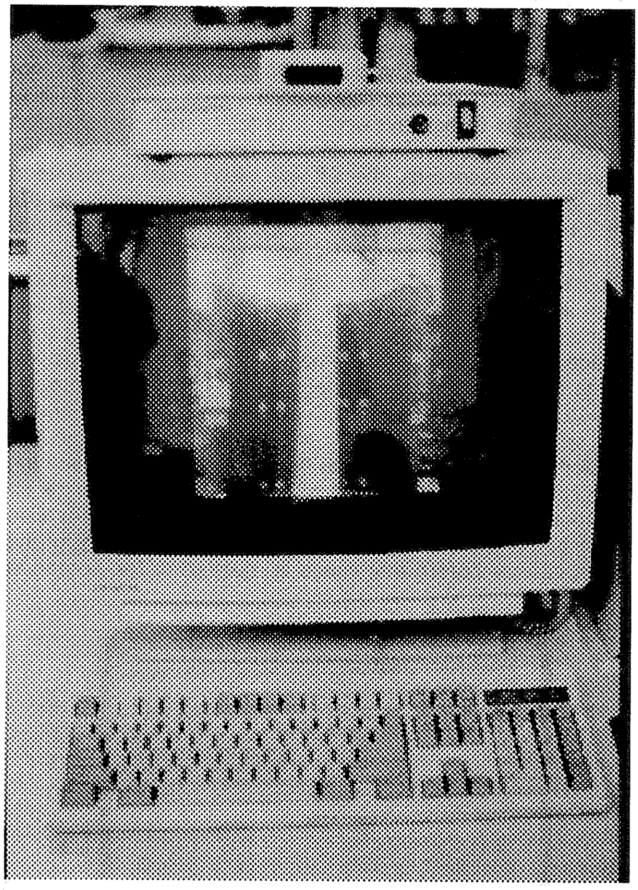

Figure 3

2.4 Interactive 3-D coordinate masurements

3-D coordinate measurements may be performed interactively on a displayed stereoscopic view by using a pointing device able to indicate positions in a virtual 3-D world. In this study a regular mouse was used and a 3-D cursor simulation was created with the help of suitable software.

A 3-D mouse cursor, or a 3-D photogrammetric floating mark in this case, consists of two separate cursors, i.e. floating mark images, one for the left and one for the right eye. Since the stereo pair has been rectified to the normal case, the two cursors must have the same y coordinate when the system is passed in stereo mode. A difference in x coordinates introducing an x-parallax may exist between the two cursors. When the mouse is used normally, the left and right motions correspond to the x coordinate and the forward and backward motions to the y coordinate. If the x-parallax is always the same, the 3D cursor is moved on the same plane (Lipton 1991). If the x-parallax increases or decreases interactively, the cursor is moved on different virtual planes. In the system under development, pressing the right mouse button indicates to the software that the distance in x direction of the two cursors should be increased, while pressing the middle button indicates that the distance should be decreased. In this way virtual motion in z is achieved.

Since the use of 3-D cursors is not directly supported by most Microsoft compatible mouse drivers, specialised software ought to be developed in order to simulate a 3-D cursor. The default mouse cursor, provided in graphics modes, is initially disabled. However, although the cursor is not visible, the mouse is working and generating interrupts to the host CPU. A new mouse cursor can be created by reprogramming the service routine which is executed when a mouse movement interrupt occurs (Gradecki 1994). The horizontal and vertical limits of the mouse movements are set equal to the left image viewport in order to ensure that the screen coordinates returned by the mouse will not be falling outside of the area used for the left image. The left cursor is mapped at a position x_L, y_L taken from the mouse movement and the right cursor is mapped at a position x_R, y_R which is calculated according to the following equations (Akka 1991b):

$$\begin{aligned} x_R &= x_L + Xparallax \\ y_R &= y_L + YOFFSET \end{aligned} \qquad (7)$$

where Xparallax is the current parallax in x direction and YOFFSET the vertical distance between the two cursors in order to have the same y coordinate in stereo mode. The distance YOFFSET depends on the current screen resolution XWIDTH, YHEIGHT and on the blank interval YBLANK inserted between the left and right image (Figure 2):

$$YOFFSET = \frac{YHEIGHT - YBLANK}{2} + YBLANK \qquad (8)$$

The exact location of the two cursors can be estimated by counting the pixels determining the Xparallax. Additional information, taken from the absolute orientation of the images, is used in order to translate x-parallax measurements to accurate Z positions in space.

The minimum configuration for the smooth operation of this system is a PC with at least a 486/100MHz processor, a 1GB hard disk drive, a mouse, a monitor capable of at least 120 Hz refresh rate and the Stereographics Corp. CrystalEyes hardware.

3. SYSTEM APPLICATION

The described system was developed for the needs of a major European Union research project, whose main objective is to examine the bahaviour of ancient monuments to earthquakes. This is a significant contribution to the protection of cultural heritage from natural hazards, especially in countries prone to earthquakes like Greece.

For the above purpose models of ancient monument parts are subjected to seismic tests using the 6DOF earthquake simulator of the Laboratory of Earthquake Engineering of NTUA. Since the movement in space of the test objects would be rather unpredictable continuous monitoring was required. Hence stereoscopic video was employed and the system described has been developed. The details and the first results of these experiments have been presented elsewhere (Georgopoulos et al. 1995). A brief description will be given here in order to present the applicability of the system.

The initial tests involved research into all aspects of the experiments. Namely the video cameras to be used, the construction of the whole setup and the algorithms employed in order to achieve the best possible results. The first simulation experiment was taped using two commercial camcorders. One was a SONY Handycam 10x Video 8 AF with 310.000 pixels and the other a SONY Handycam Pro Video 8 AF (V90) with 440.000 pixels. Around the object and independently thereof a test field was setup with 16 premarked targets. The targets were black circles on white background, approximately 30 mm in diameter. In addition three targets were attached on the object itself, in order to enable the determination of its displacements. The co-ordinates of the targets were determined using a Leica T1010 electronic theodolite equipped with a Leica DIOR 3002S EDM, with an accuracy of ±3 mm. The two cameras were positioned on tripods at a distance of 5 m from the setup with a base of 0.70 m, thus providing a base-to-distance ratio of 1:7. Imagery was acquired simultaneously with both cameras and was recorded on VHS 8mm tapes. The frames were grabbed using the Screen Machine II from FAST Electronics frame grabber with a resolution of 736(H)x560(V) pixels. This commercially available frame grabber has the usual standard features and is escorted with an image editing software with rather limited capabilities.

In order to overcome certain algorithmic problems involving two different cameras into the calculations, a second series of experiments was conducted with two identical Panasonic K900 VHS video cameras. The same targets were used and they were measured with greater accuracy using only electronic theodolite measurements from two stations. Their co-ordinates were determined with the help of intersection in space with a final accuracy of ±1 mm. In another experiment (Figure 3), where a 1:3 model of a Parthenon pillar consisting of eleven cylindrical rings was tested two identical professional Beta video cameras were used. These cameras were offered by a major Greek commercial TV Channel and were genlocked specifically for this experiment in order to produce absolutely synchronised frames. They were able to record 50 frames per second, thus increasing the sampling rate of the moving object and the amount of the data stored.

Several algorithms were either developed or used from previous work. These algorithms involved mainly a specially developed target location algorithm, a suitable camera calibration procedure and the necessary calculations for determining the ground co-ordinates of the points of interest. There were three presignalised points on each pillar ring.

By calculating the co-ordinates of these targets in space the absolute position of each one of the object blocks would be determined. This determination was carried out both monoscopically, for the case where the object was forced to swing in one plane, and stereoscopically.

In cases the pillar would perform displacements in 3D space, the object co-ordinates of the observed points are calculated by a two camera triangulation algorithm. Image co-ordinate measurements of the points of interest are performed automatically on simultaneous views of the object in order to avoid manual identification of the targets on each frame. Simultaneity is achieved by matching the frames with the same time clock indication. Since absolute synchronisation of the two cameras in the initial experiments required special hardware, which was not available, the synchronised frames were manually determined. Although the proposed method is very simple, many problems may occur in practice. The most important problem is that the clock is not clearly recorded in all video frames. Hence, due to image deterioration in certain cases it was necessary to pick frame pairs with frames having a time difference of as much as 0.05 sec. This was, however, considered as having little effect in the final result. Of course there was no synchronisation problem the case of the experiment taped with the two genlocked cameras.

The image co-ordinates of the observed pillar points are introduced into the computation as tie points with unknown positions. For each tie point 3 additional unknowns and 4 observation equations are added.

A RMS error from the comparison between the object points' co-ordinates and the co-ordinates computed by the triangulation was calculated. Concerning the first experiment with the two SONY cameras, the observed errors were 4 mm in the X and Y directions and 8 mm in the Z direction. The results are better in the case of the two Panasonic K900 cameras, where the observed errors were 1.5 mm in the X and Y directions and 4 mm in the Z direction.

These results are quite encouraging considering the hardware used. Moreover the amelioration of the final accuracies in the case of the second experiment is not a result of the different cameras used, but of the better accuracy of the measured control points' co-ordinates.

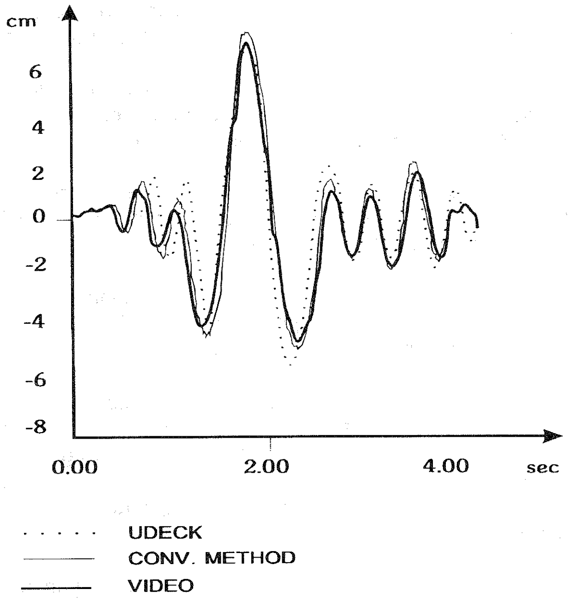

Figure 4

After the calculation of the co-ordinates of the pillar points for the whole duration of the experiments a comparison was attempted in order to assess the relative and absolute precision of the method. The comparison was performed with the similar results obtained from the conventional displacement meters measurements of the same experiment. Moreover the

results were also compared with the predicted or
with the UDECK earthquake simulation pr
diagrams were produced independently. Th
scaled to the same time and displacement
presented in Figure 4.

Similar, or even better results are expected fro calculations of the third experiment using the genlocked Beta professional video cameras.

4. FUTURE OUTLOOK

It is expected that in the near future the above described system will be enhanced with certain features and improvements, which are considered as absolutely necessary.

Firstly a better determination of the targets will be introduced. Given the limited resolutions of commercial video cameras, special measuring techniques should be used to achieve better accuracy. These techniques make use of a circular target which shows up as a rotated ellipse on the digital image. Sub-pixel edge measurements, determined via a single dimensional edge locator based on moments, are used to fit a rotated ellipse to the target (Cosandier *et al.*, 1992). Software tools for sub-pixel edge location are under development. Accuracy of 1/20th of a pixel and better is possible by using this method.

Presignalised targets should be automatically recognised, detected and measured in a fully opereational system.

The algorithm used should further be developed, in order to include determination of the reliability of the results. This would mean introducing appropriate a priori weights for the various parameters involved and calculation of the a posteriori σ_i later.

Moreover a high definition frame grabber should be available, in order to prevent eventual loss of qualitative and quantitative information during the digitization of the images.

The use of a graphics co-processor will enable the processing of more frame pairs per second, thus achieving smoother 3-D animation of the stereoscopic video.

Finally a friendly user interface should be developed to make the system amiable by non-photogrammetrists, in order for them to perform the appropriate measurements.

Furthermore, the system may later be integrated with a dedicated data acquisition system consisting of two synchronised high definition camcorders with known geometric and radiometric characteristics. The use of digital video is also an option under consideration, given the rapid development of similar commercial systems.

Application fields of the system developed are mainly industrial mensuration problems such as quality control cases, crash tests or even military implementations. Further possibilities may include kinematic studies in the field of biostereometrics.

5. REFERENCES

1. R. Akka, 1991a. Software Concepts for Stereoscopic 3D Graphics. StereoGraphics Corporation.

2. R. Akka, 1991b. Automatic software control of display parameters for stereoscopic graphics images. StereoGraphics Corporation

, T., Streilein, A., 1992. Metric accuracy
-state camera systems. *Videometrics II*,
s. 1992, pp. 103-110.

hapman, M.A., 1992. High precision target
trial metrology. *Videometrics II*, SPIE vol.
, pp. 111-122.

5. Chen, C., Scarpace F. L., 1990. One Dimensional Search for Digital Matching. International Archives of Photogrammetry and remote Sensing, Vol.28, Part 3/2. Commission III Symposium, Wuhan.

6. Gerorgopoulos, A., Tournas, E., Mouzakis, H.P., Vougioukas, E., Carydis, P. 1995. Determination of Seismic movements of monuments using stereoscopic video. Optical 3-D Measurement Techniques III, Vienna October 1995, pp.454-462

7. J. Gradecki, 1994. The Virtual Reality Constraction Kit. John Wiley & Sons, Inc.

8. He, G., Novak, K., Feng, W., 1992. Stereo camera system calibration with relative orientation constants. Videometrics II, SPIE vol. 1820, Mass. 1992, pp.2-8.

9. R. Jain, R. Kasturi, B. G. Schunck, 1995. Machine vision. McGraw-Hill

10. L. Lipton, 1991. The CrystalEyes handbook. StereoGraphics Corporation, 1991

11. Snow, W. L., Childers, B.A., Jones, S.B., 1992. Recent experiences with implementing a video based six degree of freedom measurement system for airplane models in a 20 ft diameter vertical spin tunnel. Videometrics II, SPIE vol. 1820, Mass. 1992, pp. 158-180.

12. D. Vernon, 1991. Machine vision. Prentice Hall International (UK) Ltd.

ACKNOWLEDGEMENTS

The assistance and contribution of the members of the Laboratory of Earthquake Engineering of NTUA and of MEGA Channel to this study are gratefully acknowledged.