한국측지학회지, 제 15권, 제 2호, pp.185~193(1997년 12월 30일)

Tracing of Moving Objects by Stereo Video Cameras
스테레오 비디오 카메라에 의한 운동물체의 위치추적

Lee, Chang-Kyung*
이 창 경

ABSTRACT

While close range photogrammetry has been widely applied for static deformation analysis, video cameras have many characteristics that make them the sensors of choice for dynamic analysis of rapidly changing situations. They also have limitations. The aim of this research is to explore the potential of a video system for monitoring dynamic objects.A pilot system consists of two camcorders, VCR, and PC with frame grabber. To estimte the performance of this system for moving objects, a car was imaged covering several phases when starting to drive. The sequential images of a moving car were recorded on VCR. 15 images per second were digitized in an off-line mode by frame grabber. The image coordinates of targets attached to the rear bumper of a car were acquired by IDRISI, and the object coordinates were derived based on DLT. This research suggests that home video cameras, PC, and photogrammetric principles are promising tools for monitoring of the moving objects and vibrations as well as other time dependent situations.

1. Introduction

Digital photogrammetry is rapidly emerging as a new subfield of photogrammetry. With the emergence of digital photogrammetry, expectations have been raised that the photogrammetric process may be automated. The use of video cameras and frame grabbers for non-contact measurement of physical dimensions, shape or motion is becoming increasingly popular in many fields of science and industry, including photogrammetry. Most of the research in this field has been related to three dimensional positioning. During imaging changing situations in dynamic processes, the amount of data to be manipulated with conventional analogue or analytical methods becomes prohibitively large. Recent developments in solid state cameras, image processing techniques, and microprocessors have raised the inetresting possibility of a fully automated photogrammetric system for close range applications. Compared to conventional cameras, it provides sequential images based on TV standards.

The aim of this research is to gain experience on this topic, to develop new algorithms and test existing ones, and to determine failures, their causes, and possible solutions. As a pilot system, hand-held camcorders, VCRs, frame grabber, and PC were used to take sequential images. Basic image processing software and DLT(or UNBASC2) were used to derive object coordinates. To analyze the characteristics of the camcorders, preliminary tests were conducted with moving targets. Then, to estimate the performance of this system for tracing of moving objects, a car was imaged covering several phases while moving. The images were recorded on VCR. 15 images per second were digitized in an off-line mode by frame grabber and the targets attached to the rear bumper were traced using Direct Linear Transformation(DLT).

* 군산대학교 토목공학부 조교수

2. Mathematical Model

2.1 Sub-Pixel Target Location

Precise sub-pixel target location is required for the measurements of digital images in photogrammery for control points and points of interest. This task can be accomplished by either the center of gravity method, the least-squares template matching, or the edge extraction method. Trinder's et al(1995) investigation on the precision of digital target location showed that the weighted centroid method delivers good results for flat cirular targets and is less complicated than the other two methods.

Wong and Ho(1986) developed the center of gravity formula based on the threshold window. However, target pointing by this formula was found to be subject to variations in window size, position and threshold value. Trinder(1989) added to the formula a weighting foctor for each pixel that is equal to the intensity value of the pixel above the threshold. The central high intensity pixels therefore influence the determination of the pixel location more than the surrounding low intensity pixels: i.e.,

$$c = \frac{1}{m}\sum_{1}^{n}\sum_{1}^{n} j * g_{ij} * w_{ij}$$
$$l = \frac{1}{m}\sum_{1}^{n}\sum_{1}^{n} i * g_{ij} * w_{ij} \qquad (1)$$
$$m = \sum_{1}^{n}\sum_{1}^{m} g_{ij} * w_{ij}$$

2.2 Determination of the Object Coordinates

According to Faig(1975), a non-metric camera is a camera whose interior orientation is completely or partially unknown and frequently unstable. Most of the off-the-shelf video cameras are non-metric camera. Because of the relatively large and often irregularlens distortions and image plane deformations generally associated with most non-metric cameras, the use of the analogue approach in data reduction from non-metric images is not feasible, if reasonably accurate results are desired [Karara, 1980].

Since non-metric cameras are not usually equipped with fiducial marks, a number of unique data reduction approaches not requiring the use of fiducial marks were developed. In recent years the Direct Linear Transformation [Abdel-Aziz & Karara, 1971, and Dermanis, 1994] and Self-Calibrating Bundle Adjustment [Faig, 1975b, Moniwa, 1977, and Fraser, 1982] methods have all been proposed as competing techniques for the solution of the collinearity equations in close-range photogrammetry. In this research, DLT and The University of New Brunswick Analytical Self Calibration method (UNBASC2) were selected for the determination of the object coordinates.

2.2.1 Outline of Block Adjustment

The conventional block triangulation approaches may be categorized broadly as "polynomial", "independent-model" and "bundle" approaches. In the bundle method, the basic unit is the pair of coordinates x and y of an image on the photograph. Using these coordinates, ground coordinates of interest points and estimates of the camera's orientation are derived from a simultaneous adjustment. This method differs from the sequential adjustment and independent model in that the solution leads directly to the final coordinates in a single solution and does not treat the "absolute" and "relative" orientations separately. As a result, the solution and associated error propagation are more rigorous in that certain corrections are not ignored. The procedure is iterative and requires that initial approximations be assigned to the unknown elements of camera orientation and to the initially unknown ground points.

Depending upon the degree of functional sophistication, calibration techniques are commonly classified as three basic categories: "pre-calibration", "on the job calibration" and "self-calibration" [Faig, 1975a]. The most rigorous calibration method is self-calibration. This technique differs significantly from the others in that it relies for the determination of interior orientation on the distribution of unknown object points rather than known object con-

trol points, and on the projective geometry of the multi-stereo formation overlapped by two or more photographs. In other words, these self-calibration approaches do not require object space control, except for the absolute orientation of the whole block, which requires a minimum of two horizontal and three vertical control points. Moreover, this approach has been extended to include radial symmetric, asymmetric, and tangential lens distortions. The extension of the conventional triangulation approaches, primarily developed for the bundle approach, has been accomplished by incorporating additional compensation parameters as a part of unknowns in the mathematical formulation, thus permitting simultaneous recovery of these parameters at the instant of object photography. This approach is called bundle adjustment with additional parameters or self-calibration.

2.2.2 Block Adjustment with Photo-Variant Self-Calibration

In reviewing numerous self-calibration approaches cited in the literature, it is perhaps understandable that the majority of investigations have been principally concerned with the calibration of aerial cameras, so called block-invariant self-calibration, for which a set of compensation parameters is commonly forced to remain invariant throughout the block of photographs. It has to be noted here that in some of the methods used until recently for self-calibration, the interior orientation is considered unchanged between photographs, which might not be the case for the non-metric photography, where the interior stability of the camera is rather weak.

As a photo-variant self-calibration block adjustment method, The University of New Brunswick Analytical Self Calibration Method (UNBASC2) was developed, and is based on two fundamental restraints:colinearity and coplanarity [Faig, 1975b, and Moniwa, 1977]. The collinearity equations used in UNBASC2 are the usual collinearity equations modified to include lens distortion parameters: Detailed derivation of this method is omitted here (see Moniwa 1977).

2.2.3 Direct Linear Transformation(DLT)

The innovation in the DLT approach is the concept of direct transformation from comparator coordinates into object space coordinates, thus bypassing the intermediate step of transforming image coordinates from a comparator system to a photograph coordinate system(Abdel Aziz & Karara, 1974). As such, the DLT solution makes no use of fiducial marks. The expanded method is based on the following pair of equations:

$$x+\Delta x=\frac{L_1X+L_2Y+L_3Z+L_4}{L_9X+L_{10}Y+L_{11}+1} \qquad (2)$$

$$y+\Delta y=\frac{L_5X+L_6Y+L_7Z+L_8}{L_9X+L_{10}Y+L_{11}+1}$$

$$\Delta x=(x-x_0)(K_1r^2+K_2r^4+K_3r^6+\cdots)+[r^2+2(x-x_0)^2]P_1+2(y-y_0)(x-x_0)P_2$$

$$\Delta y=(y-y_0)(K_1r^2+K_2r^4+K_3r^6+\cdots)+2(y-y_0)(x-x_0)P_1+[r^2+2(y-y_0)^2]P_2$$

where

$r_2=(x^2+y^2)$

x, y : comparator coordinates of image points

x_0, y_0 : position of the principal point in the comparator coordinates

X, Y, Z : object coordinates of the points imaged

P_1, P_2 : decentering lens distortion parameters

$L_1 .. L_2$: unknown transformation parameters

K_1, K_2, K_3 : radial lens distortion parameters

Two different approaches can be used in the formulation of the least-squares equation by means of equation (2). Each object point which has known object space coordinates(X,Y,Z) gives rise to two observation equations. If we take into accuont only K, we can derive a direct linear solution. After having obtained initial values for the unknown parameters from the linear approach, the non linear solution of equation (2) is computed by an iterative least squares adjustment.

3. Experiments

3.1 Video Image of Moving Objects

When using a non-metric camera for measurements, usually the vendor does not provide useful specifications required for data reduction processing. For that reason, the basic characteristics of the camcorder and the quality of video image had to be known, before 3 dimensional measurements were undertaken.

Basically I set up a 3 dimensional measuring system with hand-held amateur camcorders, VCRs, and a frame grabber used in the PC. They have a common drawback of low metric characteristics. On the other hand, it is easy to access and manipulate the tools, their prices are low, and the renewal interval is shorter than for the metric equipment. As an image processor, we use IDRISI for Windows. Table 1. lists the hardware and software of the system used to obtain the imagery of moving objects.

As presented in a previous paper(Lee, 1997), there were some noises in the digitized video image. However, those can be treated as acceptable random errors. In this research, we are interested in video images of the moving objects. In order to get these objectives, we designed a simple test, and conducted the experiments described below.

We know from aerial photography that the relative motion between the sensor and the object causes image blur. When it comes to imaging of moving objects, image motion depends on the speed of the moving object, the pixel size in direction of the motion, object distance, camera focal length, and shutter speed.

Theoretically, in order to get a clean image, the ratio of motion of object during exposure to the width of a TV line should not exceed 0.5. To get α practical recording of this phenomenon, we took images with specified shutter speeds, of the target board falling from different hights.

Table 1. System components

Hard/Software	Maker/Model	Specifications
Video camera I	Panasonic/AG-455P	1/3″ CCD, S-VHS, 1/2″ Tape
II	Samsung, H-33	1/3″ CCD, VHS, Hi-8 Tape
VCR	Panasonic, AG-1970	S-VHS, S-Video
Frame grabber	CreativeLab, Videoblaster SE	Resolution: 640*480(VGA)
PC	Dell, P100C	Intel P. 100 MHz 16 Mb Ram
Image processor	Clark university, IDRISI	Compatible with TiFF

3.2 3-D Measurements of a Moving Car

3.2.1 Outline of Model

The objective of this research is to find out the performance of a conventional video system, and to suggest better ways to get higher performance and accuracy in monitoring moving objects. To achieve this objective, we selected a car as α moving objects and planned to take images of the process when a driver and passenger get into the car, start the ignition, then drive ahead. We expected 3 directional movements from small to large in the process.

As shown in Fig. 1, two pairs of targets were attached to the left and right side of the rear bumper, and one pair was attached to the license plate of a car. For control, we put a 3 dimensional control frame in front of the car. The control frame has 20 control points distributed randomly, which were

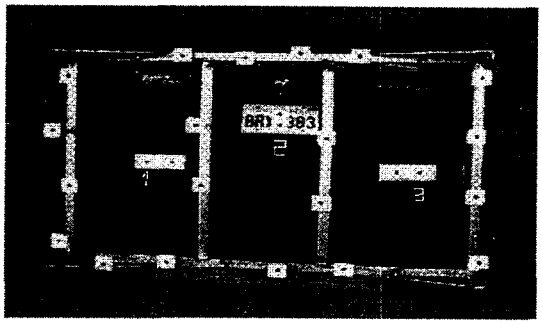

Fig. 1. Right image of a car and a control frame.

coordinated precisely with the Kern E2 Electronic Theodolite System.

3.2.2 Imaging Geometry

The adopted imaging geometry for a network is the central factor in determining the object point positioning accuracy. To obtain a near-homogeneous distribution of object space precision, a convergent multi-imaging geometry is mandatory. In the case of a series of images for a moving object, we must take into account the amount of image manipulating, therefore we used the minimum number of cameras, namely two.

An increase in the B/D ratio for near normal imaging configurations is accompanied by both an improved level of mean object point precision and enhanced reliability. In this context, we took images at three different base lengths, to compare the precision, and chose the best base length for the test.

3.2.3 Photographing Condition

Based on the preliminary tests we fixed the shutter speed at 1/1000s. Considering the field of the view of the camera(17~20°) and the size of the control frame sufficient to encompass the car, we decided on the camera stations. The diameter of circular targets that could be imaged larger than 3 pixels in x,y directions was calculated. We used black circular stickers on white background as targets. The diameter of the stickers was 19mm and the width of the background was 3 times that of the sticker's. The average distance from the camera to the objects was 2.7 m, resulting in a photo scale of 1:270. Additionally, to get a clean image of a car outdoors without artificial illumination, we needed sunshine. Unbalanced or poor illumination causes a lot of problems in target detection and location proesses.

3.2.4 Synchronization for Sequential Images

It is very important to synchronize the left and right images. Baltsavias & Stallmann(1991) used one electromic device between the camera and the VCR to add a vertical line of high contrast and binary pattern to the video signal. The binary pattern provided a code for the sequence of each image during recording and permitted safe identification of each image during the digitization. In this research, a stopwatch was attached to the front of the control frame. It refreshes its digits every 1/100s, and could be used as a code for the sequence. As an auxiliary method, the frame advancd function of the VCR can also be used for that purpose provided a significant starting point was identified in each image.

4. Test Results and Analysis

4.1 Source Sequential Imaging and Pre-Processing

The left and right video images were recorded on each S-VHS VCR. Then the VCR was connected to the PC installed frame grabber. We planned to use the stopwatch's 1/100s digits as a code for synchronization. However its letter size was too small to identify it when it was attached to the control frame located close to the rear of the car. There was no other alternative than to use the frame advancing function(1/60s) of the VCR.

Before capturing the real sequential image, significant scenes(e.g.lights on and off, door open and close) were designated as milestones. We counted the number of fields between each milestone several times. It did not exceed 2 field's(1/30s) difference during 10s. In this context, we captued every 4th field(1/15s), comprising synchronized sequential images for monitoring the movement of a car.

The PC works on a S-VGA display board, and the resolution of the images captured using the frame grabber is 788(H)*468(V). Each captured image was saved in TiFF format, its file size is 1.1Mb. This image is a 24-bit Band-Interleaved-Pixel(BIP) format, and was converted into Band Sequential(BSQ) images in the "IDRIS for Windows" format. There were no significant image quality differences between the S-VHS and VHS video cameras when the images were recorded by S-VHS VCRs using the S-Video format. 77 pairs of sequential images were digitized from S-VHS

VCR. Image coordinates of three points in each image were processed to determine their object coordinates.

4.2 Response to Moving Objects

We expected no friction when the target board was falling through the stand, but the images show that this was not quite the case. The target plate shown in Fig. 2(A) fell from 40cm height above the optical axis. Theoretically the speed should be 2.8 m/s, but its actual speed was about 1.7 m/s. The image blur was improved with the shutter speed. The relative motion during exposure was 30 rows in Fig. 2(A-1), 12 in (A-2),and 3 in (A-3). The target plate shown in the right column in Fig. 2 fell from 60cm height above the optical axis. In these series of images, because the falling speed of the target plate is faster than that for the left column, we can see more blurred images than for the left column with all shutter speeds. In static videometry, a large image scale usually provides more accurate measurements. However, in images of moving objetcts, a large image scale accompanies large relative motion in the images. Therefore the balance should be considered between measuring efficiency and relative motion. To lower the relative motion, a short time exposure is mandatory.

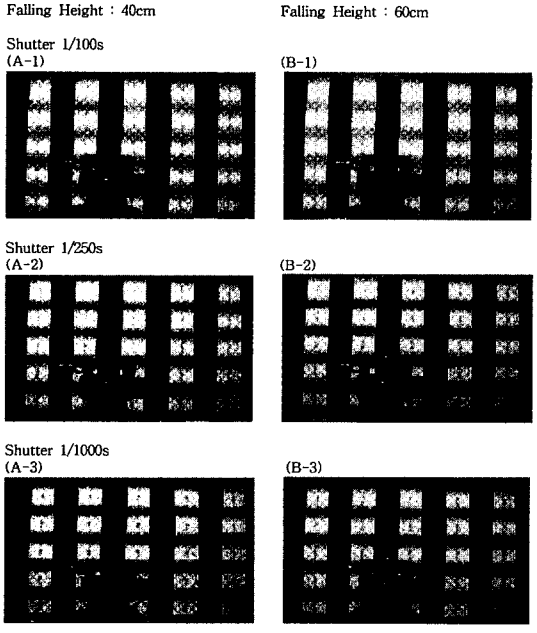

Fig. 2. Falling target images at various shutter speeds.

Usually video cameras provide a wide range of shutter speeds from 1/60 s to 1/4000s. However, it should be kept in mind that although the camera is taking a picture within a time interval shorter than 1/60s, it produces just 60 fields in a second(NTSC standard).

4.3 Sub-Pixel Target Coordinates

The image coordinates of the targets were generated from the preprocessed video image in four-step process:

* Noise removal using median filter.
* Searching window centered a target.
* Reclass based on a threshold.
* Determination of the target center coordinates using equation(1).

These peocesses for the sequential images were conducted by batch job using the macro command of IDRISI.

The enhanced images using median filter showed more condensed gray value dispersion than the original images in the histogram, and more ideal circle appearance. Target windows were detected manually, while the threshold was determined from the histogram. The change from background to target was clear in the histogram, and generally it was very close to the threshold [(maximum+mean of gray value)/2] suggested by Wong & Ho's (1986). In order to see the significance of the threshold, the target coordinates were calculated using another threshold (original threshold-4), the differences of target center coordinates obtained by each threshold were within 0.1 pixel. In Fig. 3(A) is the original target image, (B) its filtered image, and (C) its reclassed image. It was known that image enhancement causes target shift to some extent. In this test, the mean difference between the two

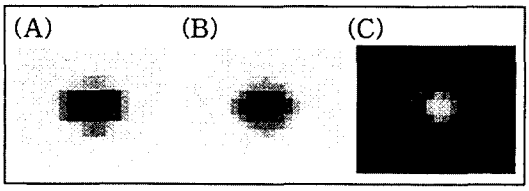

Fig. 3. Images of circular target.

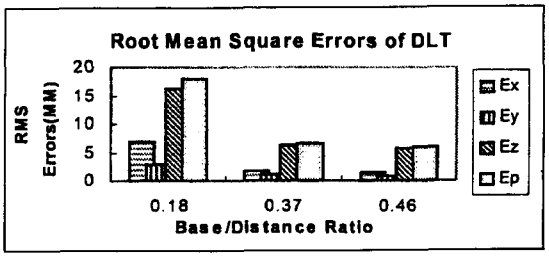

Fig. 4. RMS errors of DLT for different B/D(Unknown parameters: 12)

was nearly zero, and the maximum difference was 0.02 pixel in 20 targets.

4.4 3-D Object Coordinates

For 3-dimensional tracking of the targets in each image sequence, we prepared image coordinates for all control points and targets in the first image. Under the assumption that the control frame was fixed, the target coordinates were only prepared in the remaining image sequences. This tracking scheme reduces the work from obtaining image coordinates to running the program. As described in section 2-2, DLT and UNBASC2 could be used to derive 3-dimensional object coordinates using non-metric image coordinates. While DLT affords pixel unit as input data, UNBASC2 requires metric units. Pixel units may be considered as comparator coordinates having different scales for each axis, but UNBASC2 failed to obtain the object coordinates. Therefore, the pixel sizes of the two cameras were derived by trial and error. UNBASC2 then provided more accurate results than DLT in the test image. However, DLT was used for this experiment because it can cope with sequential data that may recourse to the same position in another sequence when the above tracking scheme was adopted.

It is well known that imaging geometry plays a great role in determining accuray. When imaging geometry is ill conditioned, the accuracy of DLT becomes worse when increasing the number of parameters. Fig. 4 shows the average root mean square errors of DLT(12 unknown parameters) for different B/D ratios. Just 20 control points were used as image coordinates. In the adopted coordinate system, the base line direction is the X axis, the azimuth direction is the Y axis, the optical axis direction is the Z axis. The error decreased when the B/D ratio increased, so we chose the images photographed at 0.46 B/D. In order to construct an optimum control set, one control point, which showed the largest RMS error was sequentially excluded from the control points. At last, optimum control points were obtained except for 2 points.

This optimum control point set provides a little smaller RMS error with the unknown parameters of DLT increaed. However, the RMS errors of the unknown points were minimized when the unknwn parameters were 12. Therefore, 12 parameters were adopted to determin the tracking of targets attached to the moving car. As shown in Table 2, RMS errors in z direction are always larger than for the other two directions.

Especially, target 2 provides smaller RMS errors than the control point set in all directions, and its magnitudes in x and y directions are within 1 mm. Generally, the accuracy between rows(y axis) is much better than that of columns in the video images. RMS errors in x direction of targets 1 and 2

Table 2. RMS errors of object coordinates(Unknown parameters of DLT: 12)

	Ex(mm)	Ey(mm)	Ez(mm)	Exyz(mm)
Control points	1.4	1.0	4.9	5.2
Target 1	14.8	10.0	73.2	75.3
Target 2	0.4	0.9	1.7	2.0
Target 3	16.2	4.8	37.4	41.2

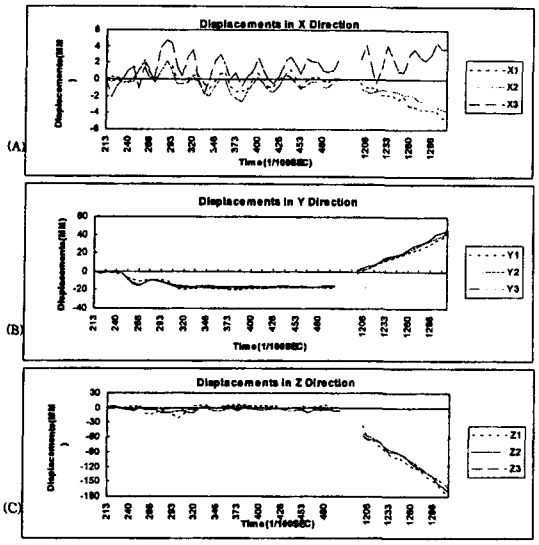

Fig. 5. Trajectory of targets attached to a moving car.

are about 1.5-3 times those in y direction. But target 2 shows a reverse phenomenon.

At last, we got the trajectory of targets attached to the car. In Fig. 5(A), we can see the vibration of the car in positive x direction owing to the impacts of shutting the driver's door, and the vibration in negative x direction owing to shutting the passenger door. After these two impacts compounded, it gradually weakened. The second half of the diagram describes the trajectory when the car starts move ahead. Target 3 was attached to a little rounded surface, and its narrow intersection angle between the two cameras might cause some large deviation of the trafectory compared to the other two targets. In y direction, the settling of the car body when driver and passenger get into the car, rebound, and final equilibrium state are recorded very cleary. In z direction, it remains still while the driver gets into the car and starts ignition. When the car starts to drive ahead, the targets' trajectories grow farther away.

5. Conculsions

Two video cameras and VCR combined with frame grabber were investigated and tested for tracing of moving objrcts. The quality of digital images photographed with the video camera, recorded on S-VHS VCR using S-Video, and digitized with the frame grabber is good enough to monitor these dynamic phenomena. In order to secure a blurless image of a moving object, it is necessary to take into account the imaging geometry, shutter speed, and illumination. Although the frame advance function of the VCR was satisfactory for synchronizstion in this experiment, more elaborate devices are indispensable to monitor high frequency vibrations. Image processing software for the PC(e.g. IDRISI) was a usable tool for image preprocessing and target centering. For real time photogrammerty system, its macro command must be improved to manipulate Windows without manual assistance. DLT was an appropriate model for the non-metric imagery except its accuracy is lower than for the self-calibration model. In conclusion, a video system and appropriate photogrammetric principles provided the expected results for monitoring of a moving car.

Acknowledgements

The authors wish to express thanks to Dr.W. Faig, University of New Brunswick, Canada, for his advices. The support of the Korean Science and Engineering Foundation is gratefully acknowledged.

References

1. Abodel-Aziz, Y.I. and Karara, H.M., "Photogrammetric Potentials of Non-metric Cameras" Civil Engineering Studies, Photogrammetry Series No. 36, University of Illinois, Urbana, 1974.
2. Baltsavias, E.P. & Stallmann, D., "Trinocular Vision for Automatic and Robust Three-Dimensional Determination of the Trajectories of Moving Objects" Photogrammetric Engineeing & Remote Sensing, 57(8), pp. 1079-1086, 1991.
3. Clarke, T.A., "An Analysis of the Prospects for Digital

Close Range Photogrammetry" ISPRS Journal of Photogrammetry and Remote Sensing, 50(3) : 4-7, 1995.

4. Dermanis, A., "Free Network Solutions with the DLT Method" ISPRS Journal of Photogrammetry and Remote Sensing, 49(2) : 2-12, 1994.
5. Faig, W. "Photogrammetric Equipment Systems with Non-metric Cameras" Close-Range Photogrammetric Systems, American Society of Photogrammetry, Falls Church, Virginia, pp.648-57, 1975a.
6. Faig, W. "Calibration of Close-Range Photogrammetric Systems : Mathematical Formulation" Photogrammetric Engineering & Remote Sensing, 41(12), 1479-1486, 1975b.
7. Fraser, C.S. "On the Use of Nonmetric Cameras in Analytical Close-Range Photogrammetry" The Canadian Surveyor, 36(3), 259-279, 1982.
8. Karara, H. M. "Non-Metric Cameras" Development in Close Range Photogrammetry-1(Edited by Atkinson), Applied Science Publishers, pp.63-80, 1980.
9. Lee, C.K., "Radiometric Characteristics of Video Images for Real-Time Photogrammetric Systems" KSCE Journal of Civil Engineering, Vol.1, No.1, pp.67-78, 1997.
10. Laurin, D. G., "A Videometric System for a Flexible Space Structure Emulator" ISPRS Journal of Photogrammetry and Remote Sensing, 48(4) : 2-11, 1993.
11. Moniwa, H., "Analytical Photogrammetric System with Self-Calibration and its Applications" Ph.D. Dissertation, University of New Brunswick, Fredericton, NB, Canada. 1997.
12. Trinder, J.C., Jansa, J. and Y. Huang., "An Assessment of Precision and Accuracy of Methods of Digital Target Location" ISPRS Journal of Photogrammetry and Remote sensing, 50(2) : 12-20, 1995.
13. Vlcek, J., "Nature of Video Images" First Workshop on Videography, ASPRS, Terre Haute, Indiana, pp.5-12, 1998.
14. Wong, K. W. and Ho, Wei-Hsin., "Close-Range Mapping with a Solid State Camera" Photogrammetric Engineering & Remote Sensing, 52(1), pp.67-74, 1986.