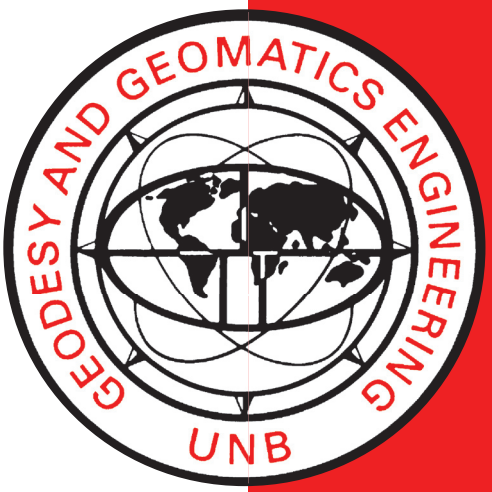
DYNAMIC MONITORING WITH VIDEO SYSTEMS

CHANG-KYUNG LEE

August 1996



TECHNICAL REPORT
NO. 183

DYNAMIC MONITORING WITH VIDEO SYSTEMS

Chang-Kyung Lee

Department of Geodesy and Geomatics Engineering
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada
E3B 5A3

August 1996

PREFACE

In order to make our extensive series of technical reports more readily available, we have scanned the old master copies and produced electronic versions in Portable Document Format. The quality of the images varies depending on the quality of the originals. The images have not been converted to searchable text.

PREFACE

This unedited technical report summarizes the research that was conducted during the author's stay as a Postdoctoral Fellow at the Department of Geodesy and Geomatics Engineering, University of New Brunswick. The research was supervised by Dr. Wolfgang Faig, and funding was provided by the Korean Science and Engineering Foundation and the Natural Sciences and Engineering Research Council of Canada.

As with any copyrighted material, permission to reprint or quote extensively from this report must be received from the author. The citation to this work should appear as follows:

Lee, Chang-Kyung (1996). *Dynamic Monitoring with Video Systems.* Department of Geodesy and Geomatics Engineering Technical Report No. 183, University of New Brunswick, Fredericton, New Brunswick, Canada, 98 pp.

Dr. Lee can be contacted at

Department of Ocean Engineering
Kusan National University
Miryong, Kunsan, Chonbuk, 573-360
South Korea

e-mail: leeck@knusun1.kunsan.ac.kr

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.

This study explores the potential of a video system for monitoring dynamic objects. The system consists of two camcorders, VCR, and PC with frame grabber. To analyze the characteristics of the camcorders, preliminary tests were conducted with still and moving targets. Then, to estimate the performance of this system for vibrations, 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 study suggests that home video cameras, PC, and photogrammetric principles are promising tools for monitoring of moving objects and vibrations as well as other time dependent situations.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my sincere thanks to Dr. W. Faig, my supervisor, for his advice, encouragement and hospitality.

I would also like to thank to John Webster, Director, Audio-Visual Services, UNB, for giving permission to use his video system, and to Hosam EL-habrouk and Xiaopeng Li, Department of Geodesy and Geomatics Engineering, UNB, for their thoughtful assistance during the experiments.

The support of the Korean Science and Engineering Foundation and of the Natural Science and Engineering Research Council of Canada is gratefully acknowledged. The author also wishes to thank the Kunsan National University in Korea for providing me with this study leave.

1. INTRODUCTION

It is well known that photogrammetry is ideally suited for precision non-contact measurements for industrial applications, particularly for objects which are difficult to be precisely measured by a direct or contact method. Most of the research in this field has been related to three dimensional positioning. When 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 interesting possibility of a fully automated photogrammetric system for close range applications. Digital cameras provide for the possibility to process the data practically simultaneously with the recording of the images. The requirement for very fast data acquisition and processing is generally dictated by the facts that certain measurement problems could not be solved otherwise and/or that the replacement of human worker by an automatic system(robot) turns out more precise and reliable, faster and less expensive results. 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, computer vision, robotics and biomechanics. Compared to conventional cameras, it provides sequential images based on TV standards.

The final objective of this research is to develop a fully automated real time photogrammetric system capable of monitoring dynamic processes. In order to reach high accuracy, the system requires sensors capable of taking images in short time intervals, data storage devices handling large amounts of data with high speed, and image processing tools that analyze the interesting features efficiently. Precise target location methods and positioning algorithms with suitable calibration techniques are a prerequisite for meeting the accuracy requirements. In the past few years, some photogrammetrists have attempted to

tackle the above mentioned problems[Baltsavias & Stallmann, 1991, Laurin, 1993, and Beyer, 1995].

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 remedies. As a pilot system, hand-held camcorders, VCRs, frame grabber, and PC were used to take sequential images. Basic image processing software as well as DLT and UNBASC2 were available to derive object coordinates.

2. CHARACTERISTICS OF VIDEO IMAGES

As a sensor, a video camera is an image plane scanner, unlike the line and "push broom" scanners that operate as object plane scanners. This means that a 2 dimensional image of the object space, defined by the field of view of the camera, is formed on the imaging surface of the tube or a solid state array. This "latent" image, scanned by an electron beam(tube camera) or the photovoltaic seats of the sensor array(solid state cameras) are "read-out" to generate a video signal, an amplitude modulated voltage signal. The video signal is then formatted to produce a 2 dimensional raster image by means of accurately supplied synchronization pulses[Vlcek, 1988].

2-1. Types of Video Cameras

2-1-1. Tube Type Sensors

The vidicon tube was the most popular and cost effective tube sensor. It employs a photoconductive layer that develops an electronic charge in response to impinging photons. An electron beam scans the photo layer in a raster format, reducing the charge along the line of the picture. This discharge produces a continues analogue signal proportional to the light intensity of the focused image. As it is seen in Fig. 2-1, the vidicon tube has a spectral response similar to, but broader than, the human eye.

The tube size is roughly analogue to the film size in a film camera, although, due to the fixed nature of the video scanning process (a constant number of scan lines per frame), a larger video tube will not necessarily provide a proportional increase in spatial resolution. Camera resolution is defined in terms of "TV lines". Lines of resolution correspond to the maximum number of

alternating white and black lines per frame height or width that can be resolved by visual inspection[Dunbar, 1986]. Tube camera resolution is a function of the electron-beam diameter relative to the area of the photoconductive layer, and generally higher than that of solid state cameras. Another advantage of vidicon tubes are that they require lower light levels than a Charge-Coupled Device(CCD). On the other hand, they require more power, are less rugged, weigh more, and are larger than solid-state devices. In addition, they exhibit significant image lag. Thus, a quickly moving bright object will appear to leave a tail in its wake, rather like a comet. It is the analog nature of tube sensors that imposes spatial and photometric accuracy limits that can be overcome by the solid-state sensors.

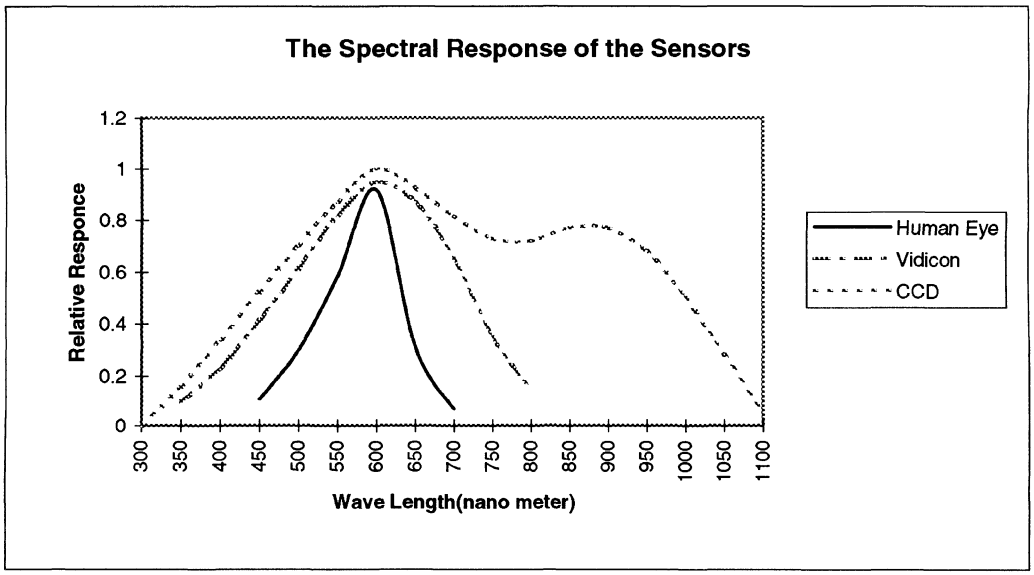


Fig. 2-1. The Spectral Response of the Human Eye, Vidicon, and CCD

2-1-2. Solid State Sensors

Since the early 1960s, self-scanning solid-state chips, based on silicon are used. Silicon , with energy sensitivity in the range of 400 to 1100 nms (see Fig. 2-1), is a good choice for detection in the visible spectrum. The decisive step forward was achieved by solving the addressing and charge transport scheme with solid-state technology. A two-dimensional solid-state imager(CCD sensor) consists of many regularly arranged sensor elements. Each cell has a light sensitive area, where are photons converted into electrons and accumulated to charge packets, and surrounding integrated electronics, which serves the purpose of transporting these packets to the output of the device.

CCD sensors most commonly use one of three addressing strategies: interline transfer, frame transfer, column-row transfer. The interline transfer CCD is organized into column pairs of devices. An imaging column of photosensors is adjacent to an opaque vertical shift register(see Fig. 2-2). Charge accumulates in the imaging column until the end of the integration period, then it is transferred to the opaque column. The signal then shifts vertically into a horizontal shift register that represents the picture sequentially line by line. This type of sensor is used in practically all consumer cameras.

In the frame transfer organization, the sensor consists of vertical columns of CCD shift registers divided into two zones. One zone, where charge accumulates during the integration time, is photosensitive. When integration is complete, the whole array is transferred in parallel to the opaque storage area of the second zone. Since the whole image zone is photosensitive, the frame transfer organization minimizes problems with Moire effects. Another advantage of a frame transfer CCD is that it can transmit one image while acquiring another. This provides flexibility to vary the integration period without changing the readout time.

A third type of solid-state sensor employes x-y, or column-row, addressing to transfer charge from the photosite to the output signal amplifier. The sensor

elements are addressed by selecting individual column and row electrodes [Dunbar, 1986].

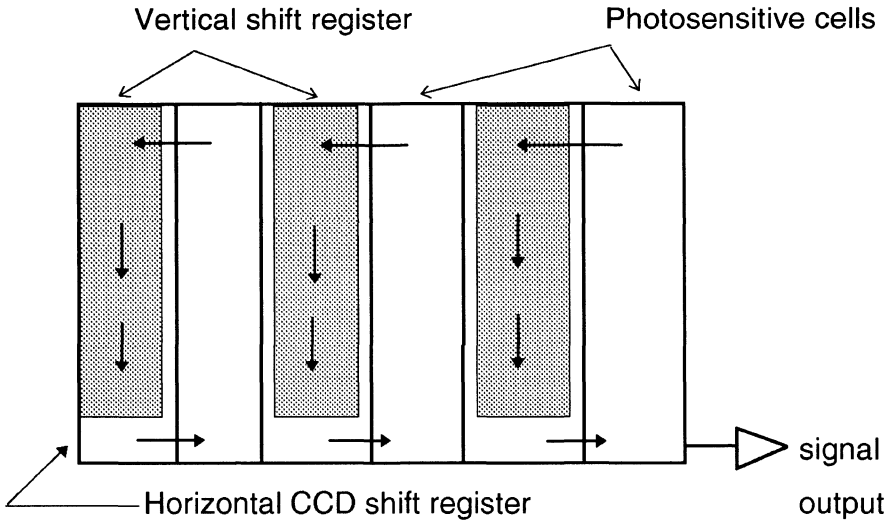


Fig. 2-2. A Diagram of an Interline transfer CCD

2-1-3. CCD Cameras

CCD cameras operate on the same basic principles as traditional cameras. The key difference between the two cameras is the media to capture an image. CCD cameras use a camera body and lens but record image data with CCDs rather than film. Fig. 2-3 shows schemitacially the optical system of a typical CCD-camera. Since optical crosstalk leads to a degradation of the modulation transfer function(MTF) at longer wave lengths for front illuminated solid-state sensors, many cameras use an infrared(IR) cut filter to eliminate light with a wavelength longer than 800nm. The diffuser is often incorporated into cameras with Interline Transfer sensor as well as into color cameras. It reduces aliasing by generating a double image displayed by ½ the sensor element pitch.

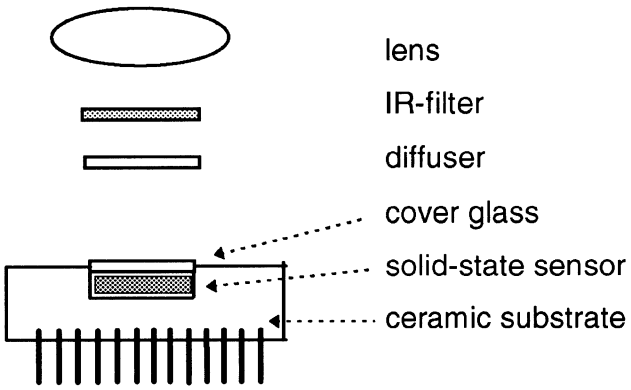


Fig. 2-3. Functional Elements of a CCD Camera

The cover-glass over the sensor serves for chemical and mechanical protection.

In a CCD camera, each of the CCD's photosensitive cells corresponds to a single pixel in the final image; the more cells the CCD provides, the higher is the potential resolution of its image. One point that complicates evaluating resolution is that the host computer may digitize the output signal at a rate inconsistent with the camera resolution.

CCD cameras offer advantages of compactness, shock resistance, low power consumption, and perfect geometry. Video cameras(e.g. camcorders) and digital cameras use these solid state sensor. Table 2-1. shows the comparison of characteristics between camcorder and digital camera.

Table 2-1. Comparison between Video and Digital Camera.

	Video Camera (Camcorder)	**Digital Camera**
Output Signal	Analogue	Digital
Scanning Method	Interlacing	Full-frame
Image Storage	Video floppy disk or cassette tape	Internal memory or memory card

2-2. Broadcasting Video Signal Standards

2-2-1. Protocols and Standards

The video image is made up of a series of horizontal scan lines, scanned left to right and from top to bottom much as a page of text is read. Within the scan, the voltage level of the signal is varied in proportion to the image brightness. The video industry uses several standards for encoding and decoding video signals. For example, the NTSC standard calls for a video frame, or image, to be scanned every 1/30s. Because this is a sufficiently slow rate for flicker to be apparent, the frame is divided into two fields, containing alternate scan lines, each displayed every 1/60s. This concept, called an interlaced signal, eliminates the flicker problem in ordinary television viewing. Twenty of the lines in each field are not used, allowing time for the scanning electron beam to be repositioned from the bottom to the top of the images. As a result, each video resolution is limited in the vertical direction to about 242.5 lines[Meisner, 1986].

In addition to the above mentioned specifications, the RS-170 standard (subset of NTSC) defines the composite video and synchronizing signal. Sync signals precede each line of video signal. The synchronization may be either internal(originating from camera) or external(originating from the frame grabber or elsewhere). The RS-170 standard specifies a 4:3 horizontal to vertical aspect ratio, thus leading to a correspondingly deformed image. On a TV monitor this distortion is automatically compensated complementary intrinsic 3:4 aspect ratio. However, the 4:3 deformed image is entering the image processing system.

In most video systems the use of television standards plays a role because the image is often transferred from camera to the frame grabber of the image processing system by an analogue video signal. These standards were developed tens of years ago for broadcast TV, long before the advent of solid state sensors and digital frame stores. The television signal can be the source of

errors and excludes the use of sensors with more rows than defined in the standard. Table 2-2. Shows the specifications of the major 3 broadcasting standards.

Table 2-2. Specifications of Broadcasting Video Signal Standards.

	NTSC (RS170)	**SECAM**	**PAL (CCIR)**
Frame Rate(Hz)	30	25	25
Scan-Lines/Frame	525	625	625
Net Scan-Lines/Frame	485	575	575
Horizontal Resolution (pixels/frame width)	250-320	416	416
Aspect Ratio(H:V)	4:3	4:3	4:3
Frequency Band used to encode the Crominance Data (MHz)	3.58	4.5	4.5
Countries of Use	USA, Canada, Japan, and Korea	France and Eastern Europe	England and the rest of Europe

2-2-2. Brightness and Color Signal Formats

Analog video signals contain two primary types of information: color and brightness. The output from different video devices combine these attributes in different formats, resulting in different levels of quality for video output. The following Table 2-3 contains the most common output formats, listed in order from lowest to highest quality

Table 2-3. Brightness and Color Signal Formats.

Signal format	Definition
Composite Video	Combines color and brightness attributes of the video signal into one signal.
S-video	Separates color and brightness into two signals to improve picture quality.
RGB	Separates the video signal into separate signals for each color components: Red, Green, and Blue(RGB).
Digital	Provides a digitized signal, using an industry-wide format, such as Recommendation CCIR601-2, or a proprietary format

2-3. Video Signal Recording

2-3-1. Principles and Formats

Video is a Latin word meaning "I see." Video recording is a form of electronic imaging whereby standard analog television signals are recorded on magnetic tape or disks.

In a video system, the CCD converts light energy into electrical signals. The raw video signal is a fluctuating DC voltage. The amplitude-modulated DC-voltage video signal is converted into a frequency-modulated signal by a video cassette recorder(VCR). Just as an audio signal can be recorded on tape, a video signal can also be tape recorded. However, recording video signals is a more difficult task because of the much higher frequencies associated with video data. High-fidelity sound recording, for instance, captures a frequency range of about 20 to 20,000Hz. In marked contrast, the frequency range of video extends to 4 MHz or more than 200 times the range of audio.

The faster the magnetic tape can be moved across the gap of the recording head, the better will be the quality of the recording of high-frequency signals. An audio tape is pulled past a stationary recording head, while a video tape is usually moved past a rapidly rotating head. Hence, the relative tape speed is dramatically increased without using a prohibitively large amount of tape.

The frequency-modulated video signal, called the luminance or Y signal, carries the black and white information of the scene. This achromatic data provides the spatial detail in the video image. The color information in the video signal, the crominance(or croma) signal, is converted by the VCR to a lower-frequency, non-modulated subcarrier. Two different color subcarrier frequencies are used, depending on the recording format[Lusch, 1988]. Table 2-4. shows various VCR formats used in broadcasting and home video systems.

Table 2-4. VCR Formats.

Format	**Maker (year)**	**Tape Width**	**Horizontal Resolution**	**Special Features**
U-Matic	Sony (1969)	3/4" (19mm)	260 lines	
VHS	JVC (late 1970s)	1/2"	240	Vertical helical scan
Beta	Sony (late 1970s)	1/2"		
Super U-Matic	Sony (1987)	3/4" (19mm)	340	
Super VHS	JVC (1987)	1/2"	420	Higher carrier frequency Improved magnetic tape Broader Y band-width Separation of Y and C
Hi-8		1/4" (8mm)	400	
ED Beta	Sony		500	Special metal-particle tape
	Sony			3 channel(RGB) recorder
	Panasonic	8" optical disk		Records 24,000 frames/disc

2-3-2. Still Frame (Freeze Frame) Function

The still frame function is important, since a single frame must be displayed during interpretation. The still frame must be free from noise bars and hold steady on the screen. Generally the VCR specifications do not make it clear whether a still frame consists of just one field, or true two field frames. A still field is preferable when image motion is present; a still frame is better if there is little image motion. Most of the frame grabbers have this function too, but some VCRs provide a still frame every 1/30s or a complete still field every 1/60s.

2-4. Frame Grabber

Since the data are processed with a digital device, it has to be either directly acquired in digital form or A/D converted. In the latter case, a separate device, fabricated by a completely different integrated circuit process than the imager, performs the conversion. Frame-grabbing refers essentially to the A/D conversion and storage of a complete image frame. Standard single circuit frame grabber boards usually come with an A/D converter, one or more frame memories, arithmetic processors, control and clocking circuits and three D/A converters for red, blue, green outputs to a color monitor from look-up tables.

The A/D converters of most frame grabbers resolve the analogue light intensity signal coming from the camera into 8 bit pixels which is equivalent to 256 shades of gray, ranging 0 for black to 256 for white or vice versa. In non-pixel-synchronized, TV-standard -based systems, the Sels(sensor elements) with the pitch s_x are read out, sampled and held with the CCD-shift register clock frequency f_s, then converted into an analogue signal with added TV line and field synchronization pulses. This analogue signal is subsequently sampled, A/D

converted and digitally stored as Pels(Picture Elements) with the clock frequency f_p of the frame grabber. Generally, in the case of Phase Locked Loop(PLL) line-synchronization, the frequencies f_s and f_p differ. Consequently Sels and of Pels are not the same size, leading to a Pel pitch:

$$p_x = s_x \frac{f_s}{f_p} \approx s_x \frac{N_{Sels}}{N_{Pels}} \quad \text{.....(2-1)}$$

in which N_{Sels} is the number of active Sels per line and N_{Pels} is the number of Pels per line [Lenz & Fritsch, 1990]. The resulting number of pixel is dependent on the number of stored elements and the frequency of the pixel clock on the frame grabber. Normally neither f_s nor f_p are known or specified with sufficient accuracy.

Another error source, influencing the accuracy of a digital image is the line jitter. It is caused by the fact that the beginning of the line, defined by the horizontal sync pulse, is not detected by the frame grabber with sufficient accuracy.

Earlier investigations have shown that synchronization errors can reach up to 0.1 pixel in a line. In order to achieve highest accuracy in photogrammetry and image processing, direct digital readout of CCD sensor is mandatory. Another way to reduce synchronization errors is external synchronization. Here the pixel clock of the frame grabber is used to synchronize the camera(or vice versa). Since the systems with standard TV-signal input are still very common, a scheme to determine line jitter, drift and p_x / s_x with high precision was proposed by Lens(1989).

If A/D conversion and/or the memory is too slow, the image will have to be stored over a number of frames by storing every n^{th} pixel per unit time. At a frame rate of 30 frames per second, it becomes necessary to cope with data sets of 640*480*8 bits in 1/30 s. No conventional PC can cope with these demands within 1/30 s, and it is necessary to introduce additional processing power into the PC. A survey of frame grabbers is given in Gruen(1988).

2-5. Digital Image Display

In photogrammetry, our traditional medium for storing information about an object of interest is the photograph. With digital photogrammetry, the medium is a digital image. The digital imagery which is collected, may need to be displayed to allow measurement and interpretation of information and also to verify results obtained automatically. For simple interpretation and verification, it may be suffice to view the digital images monoscopically, but when three dimensional geometric information is to be checked or measured, stereo viewing is necessary. At present, the options available for stereoscopic viewing are similar to those which have been used on analogue stereorestitution instruments in the past and include anaglyphic filters, polarizing systems and binoculars.

A familiar binocular viewing system may be mounted and the left and right image independently presented either by using two monitors or displaying both images on one monitor which has split screen capabilities. The anaglyphic filter principle may be applied by displaying the left and right images in a monitor using complementary colors, usually red and green. The stereo model is seen by viewing the monitor through the glasses with the appropriately colored filters covering the eyes. Lastly, a polarizing system may be effected for digital imagery by placing a liquid crystal shutter in front of the monitor with the left and right image being alternatively displayed and polarized. The user wears glasses with polarizing filters to see stereoscopically.

2-6. Radiometric Calibration of Video Images

From its generation at the sensor to its use in the computer, each pixel undergoes a number of operations, geometric and radiometric in nature, that may lead to distortions both in geometry and radiometry. When analyzing error sources and their effect on pixels it is important to keep the duality between pixel

radiometry and information location in mind. A change in pixel gray value leads to a geometric distortion of an image feature (although the physical location of the pixel itself is not altered) and vice versa. Therefore the geometric and radiometric distortions may have similar effects on the final results. Geometric calibration consists of the evaluation of the array regularity, principal point and principal distance, and lens distortions. For radiometric calibration, Curry et al. (1986) investigated detector response linearity, system noise, and system linearity over time. Generally, radiometric corrections should be applied before any geometric operations are performed.

2-6-1. Sensor Linearity

The CCD is a semiconductor device. Pixel illumination is converted to an output voltage by the array, and then to a digital gray value in the range of 0(black) to 255(white) by the analog to digital converter. Ideally, the entire conversion should be linear.

2-6-2. System Noise

System noise is defined as fluctuations in pixel gray levels caused by random and systematic perturbations in the CCD sensor and A/D converter.

Systematic noise is seen as a repeated spurious response by sensor elements. A spurious response is generally defined as an output of more than 10 percent of the saturation voltage, in comparison with the pixel's nearest neighbors. Blemished pixels are determined by running the camera in the dark(lens cap on) and then measuring the output from each pixel. There will always be a minimum signal called the dark signal, defined as the output signal caused by thermally generated electrons. Responses above this dark current are spurious, and the pixels affected must be rejected from subsequent images if the response cannot be corrected.

Random noise results from imperfections in the sensor manufacture, electron leakage between sensor photosites, variations in the A/D conversion, and other factors. Levels of random noise can be approximately quantified by running the camera at various clock rates, which affect pixel integration times, and under different lighting conditions, while viewing a neutral surface. Blooming occurs when a bright region in the frame locally overloads the sensor, causing overflow into adjacent pixel rows and columns beyond the actual bright spot.

2-6-3. Sensor Stability

After the camera and associated electronics are switched on, the sensor and controller undergo thermal and other changes. A test was devised to check the internal stability of the sensor readout over a period of eight hours. A single discrete point was placed in the field of view of the camera, and an image was stored every hour for eight hours. Using the sub-pixel location techniques, the point location in x and y was computed and compared [Curry et al., 1986].

2-7. Configurations of Video Systems

Table 2-5 shows hardware components of video systems which were configured for various photogrammetric tasks. There is a continuous trend that the size of sensor pixels become smaller, while the number of sensor pixels become larger.

Table 2-5. Configurations of Video Systems.

Researcher (Year)	Curry et al. (1986)	EL-Hakim, (1986)	Vosselman & Förstner(1988)	Ruther & Parkyn (1990)
Camera				
Output Signal	Video	Video	Video	Video(CCIR)
Maker	GE	Hitachi	Hamamatsu	Siemens
Model	TN2200	KP-120	C1000	K211
Lens		12.5-75mm	57.5mm	17mm & zoom
Sensor				
Type	CID	MOS	MOS	CCD
Maker, Model			Hitachi-He97211	Sony XC57CE
Array(H*V)	128*128	320*244	256*320	500*582
Pixel Size	45.72*45.72 μm^2	27*27 μm^2	21.6*21.6 μm^2	17*11 μm^2
FrameGrabber				
Make				Matrox
Model	PN2200			PIP-512,1024
Array(H*V)				512*512
Pixel Size				15*11 μm^2
Computer				
Maker	DUAL 83/80	Motorola		IBM PS/2 M30
Processor		MC68010		Intel80386
OS	UNIX	UNIX cousin		DOS

Researcher (Year)	Heipke et al. (1992)	Beyer (1992)	Trinder,Jansa& Huang(1995)	Lee & Faig (1996)
Camera				
Output Signal	Digital (Interline)	Video(interline)	Video	Video (NTSC)
Maker,	Kontron	Sony		Samsung
Model	ProgRes 3000	XC-77CE		H33
Lens		Fujinon 9mm	16mm	6-48mm, f/1.8
Sensor				
Type	CCD	CCD	CCD	CCD
Maker, Model				Sony,ICX 058AK
Array(H*V)	500*582		700*500	811*508
Pixel Size	17*11 μm^2		10*10 μm^2	6.35*7.4 μm^2
FrameGrabber				
Maker,		Datacube,		Creative Lab,
Model		MAX-SCAN		VB SE
Array(H*V)		726*568	512*512	788*468(SVGA)
Pixel Size		11*11 μm^2		
Computer				
Maker		SUN		IBM PC(Dell)
Processor	IBM-AT	SPARK S.1		XPS P100C
OS		UNIX		DOS,Window95

3. DIGITAL TARGET LOCATION

With the advent of digital photogrammetry, expectations have been raised that the photogrammetric process may be automated. A fundamental process is the orientation of a stereo pair. The problem of automatic orientation can easily be formulated: Select a sufficient number of points in one image and find the corresponding points in the other image, followed by an adjustment to determine the orientation elements. The crucial step is obviously to find corresponding(conjugate) points. This task is accomplished by image matching methods, such as gray level correlation or feature matching.

Precise sub-pixel target location is required in the measurements of digital images in photogrammetry for control points and interest points. Mikhail et al.(1984) have studied the detection and location of edges and cross-targets , while Wong and Ho(1986) have developed a method for the location of circular targets on a digital image and, Trinder(1989) enhanced the accuracy of the method with introducing a weight for each pixel.

On the other hand, Ackermann(1984) presented a digital image correlation algorithm based on least squares windows matching to measure parallax and point location. In this method(so-called area-based image correlation method), matches are found by correlating two images(windows), from the left and right photographs, based on raw gray level values of the pixels.

Another feature that can be used as the element to be matched is a point. In the interest operator method, distinctive pixels are evaluated rather than areas. Schenk and Hoffmann(1986) presented a method of stereo matching using edges that were detected by Laplacian of Gaussian(LoG). Trinder et al(1995) assessed the precision and accuracy of methods of digital target location.

Even though we use any of the method, the size, shape and contrast of the target directly affect the accuracy of the targeting operation.

3-1. Image Coordinate System

A pixel-based image coordinates system was chosen for the calculations with the video cameras. The location of the principal point is not specified for most CCD-cameras, varies from camera to camera and depends on the configuration of the frame grabber.

The pixel coordinates of all targets were measured by the method described in section 6-1-2. The pixel coordinates were subsequently transformed to image coordinates with the pixel-to-image coordinate transformation:

$$x = (x' - x_p') \times p_x \qquad \text{.....(3-1)}$$

$$y = (y' - y_p') \times p_y$$

where x, y : image coordinates

x', y' : pixel coordinates

x_p', y_p' : location of principal point in pixel coordinates

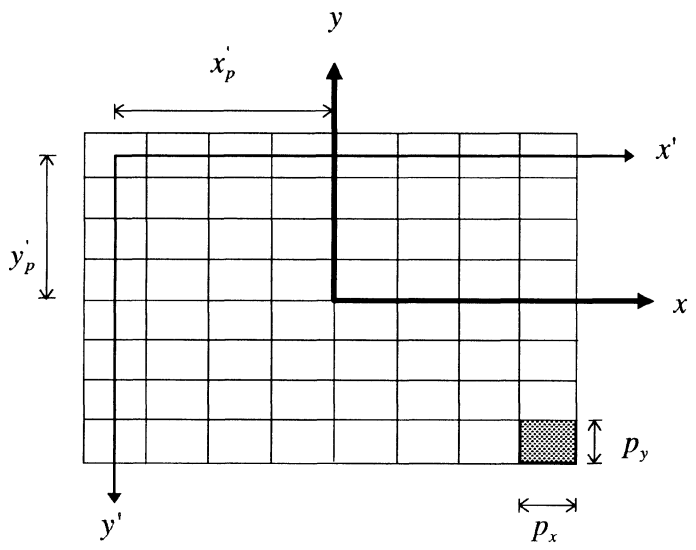


Fig. 3-1. Pixel and Image Coordinates Systems

3-2. Algorithms for Sub-Pixel Target Location

3-2-1. The Center of Gravity

A rectangular window of suitable size so that it covers the complete target and the surrounding area is approximately centered on the target. Thresholding within the window is then carried out. Thresholding converts the window area to binary values with all pixels whose intensity is above a threshold set to 1, and the remainder to zero. This process identifies the area of the target. Initially, the threshold was computed using following formula:

Threshold = (min pixel value + mean pixel value) / 2 .....(3-2)

Subsequently, this formula was modified to eliminate systematic errors caused by target asymmetry. Wong and Ho(1986) then chose to compute the position of the target by taking the center of gravity using:

$$x = 1/M \sum_1^n \sum_1^m j * g_{ij}$$

$$y = 1/M \sum_1^n \sum_1^m i * g_{ij} \qquad \text{.....(3-3)}$$

$$M = \sum_1^n \sum_1^m g_{ij}$$

where g_{ij} is the value either 1 or 0 of each pixel, located in row i column j.

Target pointing by the above formula, however, was found to be subject to variations in window size and position, as well as the threshold value, which caused variations in the position of the target of up to ½ pixel. This was because low intensity pixels on the edge of the target but still above the threshold had a disproportionately large influence on the location of the target.

It was therefore necessary to add to equation (3-3) a weighting factor w_{ij} for each pixel that is equal to the intensity value of the pixel above the threshold as shown in equation (3-2).

The central high intensity pixels therefore influence the determination of the pixel location more than the surrounding low intensity pixels[Trinder, 1989]: i.e.,

$$x = 1/M \sum_1^n \sum_1^m j * g_{ij} * w_{ij}$$

$$y = 1/M \sum_1^n \sum_1^m i * g_{ij} * w_{ij} \quad \text{.....(3-4)}$$

$$M = \sum_1^n \sum_1^m g_{ij} * w_{ij}$$

3-2-2. Method based on Least Squares Matching

The digital image correlation procedure starts from digitized homologous image areas(windows) that are given in the form of matrix arrays g(x, y) of observed gray level values. The optimum match is defined by the transformation of the left and right window of one array onto the other which minimizes the remaining gray value differences. The transformation contains geometric and radiometric corrections and the transformation parameters are obtained by an iterative least squares solution.

The given gray value arrays $\overline{g}_1(x,y)$ and $\overline{g}_2(x,y)$, respectively, represent the quantified image functions $g_1(x,y)$ and $g_2(x,y)$ which are perturbed by noise $n_1(x,y)$ and $n_2(x,y)$ respectively:

$$\overline{g}_1(x,y) = g_1(x,y) + n_1(x,y)$$

$$\overline{g}_2(x,y) = g_2(x,y) + n_2(x,y) \quad \text{.....(3-5)}$$

All functions relate to the same coordinate system, defined by the(identical) pixel arrays.

It is now assumed that $g_2(x,y)$ is a transformation of $g_1(x,y)$, as both are images of the same object. For simplicity, an affine(6 parameter) transformation is considered as a geometric transformation, and 2 linear parameters are included for the radiometric transformation :

$$g_2(x,y) = h_0 + h_1 * g_1(a_0 + xa_1 + ya_2,\ b_0 + xb_1 + yb_2) \qquad \text{.....(3-6)}$$

where

h_0, h_1 are zero value shift and brightness scale factor, and

$a_0, a_1, a_2, b_0, b_1, b_2$ represent perspective distortion and relief displacement.

Expanding equation (3-6) to linear increments from approximate values $g_2{}^0(x,y) = g_1(x,y)$, we obtain

$$g_2(x,y) = g_1(x,y) + \dot{g}_x da_0 + x\dot{g}_x da_1 + y\dot{g}_x da_2 + \dot{g}_y db_0 + x\dot{g}_y db_1 + y\dot{g}_y db_2 + dh_0 + g_1(x,y)dh_1 \qquad \text{.....(3-7)}$$

where $\dot{g}_x(x,y) = \partial g_1(x,y) / \partial x$ and $\dot{g}_y(x,y) = \partial g_1(x,y) / \partial y$ are gradient functions of $g_1(x,y)$. Considering now the difference, we get

$$\overline{\Delta g}(x,y) = \overline{g_2}(x,y) - \overline{g_1}(x,y) = g_1(x,y) - g_2(x,y) + n_2(x,y) - n_1(x,y) \qquad \text{.....(3-8)}$$

When replacing $g_2(x,y)$ by the expansion of equation (3-7), substituting $n_2(x,y) - n_1(x,y) = v(x,y)$, and switching back to discrete coordinates x_i, y_j, we obtain

$$\overline{\Delta g}(x_i, y_j) + v(x_i, y_j) = \dot{g}_x(x_i, y_j)da_0 + x_i\dot{g}_x(x_i, y_j)da_1 + y_j\dot{g}_x(x_i, y_j)da_2 + \dot{g}_y(x_i, y_j)db_0 + x_i\dot{g}_y(x_i, y_j)db_1 + y_j\dot{g}_y(x_i, y_j)db_2 + dh_0 + g_1(x_i, y_j)dh_1 \quad \text{.....(3-9)}$$

For each pixel i, j, equation (3-9) represents linearized observation equation which relates the observed gray value difference $\overline{\Delta g}(x_i, y_j)$ to the 8 unknown transformation parameters $a_0, a_1, a_2, b_0, b_1, b_2, h_0, h_1$. The least squares solution minimizes the square sum of the residuals $v(x_i, y_j)$, thus representing the desired optimum transformation with the optimum window matching[Ackermann, 1984]. However, it requires iterations, which implies resampling. Both windows are compared pixel by pixel, related by the identical pixel coordinates. After a transformation, this relation is destroyed. Therefore, after each iteration step, the gray level values have to be interpolated in the transformed coordinate system. Among the nearest neighbors, bilinear interpolation or cubic convolution is used for resampling.

The coefficients of equation (3-9) also contain the unknown image function $g_1(x_i, y_j)$ and its gradient functions $\dot{g}_x(x, y)$ and $\dot{g}_y(x, y)$. Being not directly available, they have to be estimated from the observed function $\overline{g}_1(x_i, y_j)$ by applying suitable smoothing procedures for filtering the noise $n_1(x, y)$. The solution depends very much on the gradients which represent the image texture. This approach to image correlation offers a very good estimate of $\hat{\delta}_0^2$, indicating the goodness of the match between the correlated arrays.

This correlation algorithm is non-symmetrical, so the correlation of the right to left hand window is not exactly identical with that in the reverse direction. The procedures of least squares matching is as follows.

a. In the right image, select the region, which is analogue to the reference window of the left image.

b. Design an image function for the selected window.

e.g.) image function : $g(x_i, y_j) = c_0 + c_1 x_i + c_2 y_j + c_3 x_i^2 + c_4 x_i y_j + c_5 y_j^2$

c. Derive the coefficients (c_i) of the image function by the least squares.

d. Derive $\dot{g}_x(x,y) = \partial g_1(x,y) / \partial x$ and $\dot{g}_y(x,y) = \partial g_1(x,y) / \partial y$.

e. Construct the observation equation (3-9).

f. Derive the transformation parameters of equation (3-9) by least squares.

g. Resample using the above transformation parameters $(a_0 \sim b_2)$.

h. Move search window using geometric transformation parameters.

i.e. $x^n = (a_0 + da_0) + (a_1 + da_1)x^o + (a_2 + da_2)y^o$

$y^n = (b_0 + db_0) + (b_1 + db_1)x^o + (b_2 + db_2)y^o$

i. Repeat the above procedures (a-h) until the transformation parameters become smaller than certain limits.

3-2-3. Method Based on Edge Detection

(1) Edge Detection

Edges correspond to abrupt light intensity changes in the image. The subject of determining edges is intensively discussed in computer vision, and many operators with different properties have been proposed. Schenk et al.(1991) used a LoG operator to detect edges. The Gaussian filter,

$$G(x,y) = \frac{1}{2\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \text{.....(3-10)}$$

is ideal because it optimizes the two conflicting constraints of being limited in the spatial and frequency domains. The second derivative of the image indicates abrupt intensity changes. The two operations, that are smoothing and taking the second derivative, can be combined. Taking the second derivative of the Gaussian G(x,y), we obtain the definition of the LoG operator; i.e.,

$$\nabla^2 G(x,y) = \left[\frac{x^2+y^2}{\sigma^2} - 2\right] e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad \text{.....(3-11)}$$

From equation (3-11) we conclude that the width of the central lobe w, is related to σ, the parameter of the Gaussian filter, by $w = 2\sqrt{2}\sigma$

Convulsing the image g(x,y) with the LoG, we obtain the convoluted image

$$c(x,y) = \nabla^2 G(x,y) * g(x,y) \qquad \text{.....(3-12)}$$

Edges are found at locations, called zero crossings, where c(x, y)=0. The average distance between neighboring zero crossings is approximately w.

(2) Center of Target to Sub-Pixel Accuracy

The edges found by zero crossings are consecutive pixels. If we adopted circular targets, their edges comprise circumferences. When it comes to orientations and 3 dimensional measurement of interest points, we need target points, not target areas. We must therefore derive the location of targets to sub-pixel accuracy. Here we introduce two basic methods.

① The Interpolation of the Correlation Window

Finding the precise location of points is achieved in two steps. First, one must compute the correlation between the reference target and the searching target. Then sub-pixel accuracy is computed by determining the maximum of the surface constructed by the correlation values in 3*3(any size) windows. i.e.,

$$c(x_i, y_j) = a_0 + a_1 x_i + a_2 y_j + a_3 x_i^2 + a_4 x_i y_j + a_5 y_j^2 \quad \text{.....(3-13)}$$

The coefficients a_i are obtained by least-squares adjustment of the second order two dimensional orthogonal polynomial. The sub-pixel location x_p, y_p is obtained from Equation (3-14): i.e.,

$$x_p = \frac{a_2 a_4 - 2a_1 a_5}{4a_3 a_5 - a_4^2}$$

$$y_p = \frac{2a_3 x_p + a_1}{a_4} \quad \text{.....(3-14)}$$

② The Intersection of Target Axes

A circular dot is a common type of target, however, its projection on to an image is, in most case, an ellipse. The center of the ellipse is not the center of the circular dot. Nevertheless, this is negligible for most cases such as those typically found in photogrammetry. A method of finding the center of the ellipse is to find two diameters and then calculate their intersection which is the center of the ellipse. We can find two end points of a set of these chords by scanning the ellipses along the row(or column) direction or using another edge detection operator. But in this case, we should know the boundary location to sub-pixel accuracy. Tabatabai and Mitchell(1984)'s moment preservation method has been used for sub-pixel boundary searching.

3-3. Auxiliary Validity Checks

Wong & Ho(1986) performed a validity check on each control point using criteria of shape, size, and location within the search window. In case the control points are circular in shape, the ratio of the second moments(I_x and I_y) about two principal axes should be approximately equal to 1. The following computation formulas were used:

$$I_x = \sum_1^n \sum_1^m i^2 * g_{ij}$$

$$I_y = \sum_1^n \sum_1^m j^2 * g_{ij} \qquad \text{.....(3-15)}$$

$$I_{xy} = \sum_1^n \sum_1^m i * j * g_{ij}$$

$$I_{x'} = \frac{I_x + I_y}{2} + \sqrt{(\frac{I_x - I_y}{2})^2 + I_{xy}{}^2}$$

.....(3-16)

$$I_{y'} = \frac{I_x + I_y}{2} - \sqrt{(\frac{I_x - I_y}{2})^2 + I_{xy}{}^2}$$

$$R = \frac{I_{x'}}{I_{y'}} \qquad \text{.....(3-17)}$$

If the ratio R exceeded a certain limit for a given control point image, it was rejected as a control point. Furthermore, a control point was also rejected if the smaller of the second moment($I_{y'}$) was close to zero, or if a non-zero pixel was located on the outer-most row or column within the search window.

3-4. Comparison of Algorithms

Table 3-1. Comparison between Methods of Sub-Pixel Target Location.

Photographing

Researcher	Wong & Ho	Curry et al.	EL-Hakim	Vosselman & Förstner
Year	1986	1986	1986	1988
Camera-Obj. Distance(m)	0.866	0.3, 0.338, 0.376		5
Baseline(m)	0.108			
Exposure Stations	2	1	2	9(simulated by rotations)
Parallel/Convergent	parallel			convergent

Target(control points)

Shape	circular	dots	circular	square, circle triangle, cross
Size(mm)	D=6	D=0.5	D=3	9-15, 17-25
Color/Back-ground	black/white	white/NA	B/W	black/white
Control-Survey /Accuracy	x, y : 0.2, z : 0.4(mm)	Wild A5 co-ordinatograph x, y : 0.01(mm)	SIP model 560M	

Sub-pixel Target Location

Method	Center of Gravity	2-D Parabolic Curve Fitting	Centroid and Gray-level Interpolation	Least Square Matching
Additional Checking	The Ratio of 2nd moments	1-D Parabolic Curve Fitting		
Image Enhancement		Laplacian Filter	Linear and Non-linear Convolutions	
Thresholding	(mean+min) * 0.5 + 0.99		histogram	yes
Target Size(pixels)	4*4	3*3, 4*4, 5*5		4 - 7, 8 - 12
Precision(um)/ (pixel)	2.4/0.4	/0.2		/0.1

[Continued Table 3-1.]

Photographing

Researcher	**Ruther& Parkyn**	**Beyer**	**Trinder,Jansa & Huang**	**Lee & Faig**
Year	1990	1992	1995	1996
Camera-Obj. Distance(m)				2.6
Baseline(m)				1.2
Exposure Stations		6 (with rotation 0, 90 degree)	2	2
Parallel/Convergent			Convergent	Convergent

Target(control points)

Shape	hollow cylinder	circles	Circular, Cross, Sphere	Circular
Size(mm)		20	12.7(ball)	19
Color/background		black/white		black/white
Control - Survey /Accuracy	Zeiss UMK 10 Camera	Industrial Theodolite S. /1mm	/1mm	Kern E2 Electronic Sys /0.03mm

Sub-pixel Target Location

Method	Weighted Mean	Least Square Matching	Edge of Target	weighted Gravity Center
Additional Method			Template-Mat. W. Center	
Image Enhancement	yes		Gaussian Filtering	median filter
Thresholding	yes		yes	histogram
Target Size(pixels)		2.4 - 4.0	3 - 11	10-16
Precision(um)/ (pixel)	1/0.08		/0.004-0.02	/0.1

4. THREE DIMENSIONAL OBJECT COORDINATES

According to Faig(1975a), 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 belong to non-metric camera. Because of the relatively large and often irregular lens 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, Bopp & Krauss, 1977, Faig & Shih, 1986, Naftel & Boot, 1991 and Dermanis, 1994] and Self-Calibrating Bundle Adjustment [Faig, 1975b, Moniwa, 1977, Granshaw, 1980 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.

4-1. Block Adjustment with Photo-Variant Self-Calibration

4-1-1. Outline of the 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]. A most rigorous calibration in this context can be accomplished with the method of 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 control 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.

Depending upon the practical considerations of introducing additional parameters, systematic image distortion has been parameterized with many variations, which can be divided into two major categories. The first is concerned with the "decomposition" of image distortion into various components of lens distortion and film deformation. This procedure is based on the assumption that all factors affecting image perturbations could be physically interpreted, thereby

predicting an individual mathematical model for specific distortion components. A somewhat more complex situation exists when treating the sophisticated distortion characteristics that are not explicitly interpretable, and hence the composite effects of systematic errors are modeled as an entity regardless of the contribution from each individual source. This approach forms the basis of the second category of parameterization technique.

In reviewing numerous self-calibration approaches cited in the literature, it is perhaps understandable that the majority of investigations has 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. The shortcomings raised by block-invariant self-calibration can be offset if the independent set of compensation parameters is postulated for each individual photograph involved. Although this approach requires a more laborious computational effort, it is a generalized self-calibration scheme applicable for any type of camera (metric or non-metric) and photography (aerial or close-range).

4-1-2. The University of New Brunswick Analytical Self Calibration Method (UNBASC2)

The photo-variant self-calibration block adjustment method(UNBASC2) was developed at UNB, 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:

$$F_{xij} \equiv \left[\left(x_{ij} - x_{0j}\right) + \Delta x_{ij}\right] \bullet M_{3j} + C_j \bullet M_{1j} = 0$$

$$F_{yij} \equiv \left[\left(y_{ij} - y_{0j}\right) + \Delta y_{ij}\right] \bullet M_{3j} + C_j \bullet M_{2j} = 0 \qquad \text{.....(4-1)}$$

where $\left[M_1, M_2, M_3\right]_j^T = [R]_j \bullet \left[\left(X_i - X_{cj}\right), \left(Y_i - Y_{cj}\right), \left(Z_i - Z_{cj}\right)\right]^T$

$(X,Y,Z)_i$: unknown ground coordinates of object point i,

$\left(X_c, Y_c, Z_c\right)_j$: unknown ground coordinates of the exposure station of photograph j,

$[R]_j$: the 3*3 rotation matrix of photograph j in terms of rotations $\omega, \phi \; and \; \kappa$

$(x,y)_{ij}$: comparator coordinates of object point i measured on photograph j,

$\left(x_0, y_0, c\right)_j$: unknown parameters of principal point and principal distance of photograph j,

$\left(\Delta x, \Delta y\right)_{ij}$: the x and y components of image distortion of object point I on photograph j.

In equation(4-1), the image distortion function involves the conventional distortion functions of radial-symmetric and decentering lens distortions, as well as film shrinkage and non-perpendicularity of comparator axes. By introducing the latter, the image coordinates are solely based on the comparator system, which makes the need for fiducial marks optional rather than mandatory. The distortion parameters are defined as:

$$\Delta x_{ij} = dr_{xij} + dp_{xij} + dq_{xij}$$

$$\Delta y_{ij} = dr_{yij} + dp_{yij} + dq_{yij} \qquad \text{.....(4-2)}$$

where $dr_{xij} = (x_{ij} - x_{0j})(k_{1j} \bullet r_{ij}^2 + k_{2j} \bullet r_{ij}^4 + k_{3j} \bullet r_{ij}^6)$

$$dp_{xij} = p_{1j} \bullet [r_{ij}^2 + 2(x_{ij} - x_{0j})^2] + 2p_{2j}(x_{ij} - x_{oj})(y_{ij} - y_{0j})$$

$$dq_{xij} = A_j \bullet (y_{ij} - y_{0j})$$

$$dr_{yij} = (y_{ij} - y_{0j})(k_{1j} \bullet r_{ij}^2 + k_{2j} \bullet r_{ij}^4 + k_{3j} \bullet r_{ij}^6)$$

$$dp_{yij} = p_{2j} \bullet [r_{ij}^2 + 2(y_{ij} - y_{0j})^2] + 2p_{1j}(x_{ij} - x_{oj})(y_{ij} - y_{0j})$$

$$dq_{yij} = B_j \bullet (y_{ij} - y_{0j})$$

$$r_{ij}^2 = (x_{ij} - x_{0j})^2 + (y_{ij} - y_{0j})^2$$

and $(k_1, k_2, k_3)_j$ are the parameters for radial symmetric lens distortion of photograph j,

$(p_1, p_2)_j$ are the parameters for decentering(asymmetric radial and tangential) distortion of photograph j,

$(A, B)_j$ are the parameters for scale differences along the comparator axes and their possible non-perpendicularity on photograph j,

from which it is possible to interpret some of the distortion characteristics by the following relations:

$$P = r^2 \bullet \sqrt{[p_1^2 + p_2^2]} \quad \text{.....(4-3)}$$

$$\theta = \tan^{-1}[-p_1 / p_2] \quad \text{.....(4-4)}$$

$$\lambda_y = \sqrt{[A^2 + (B+1)^2]} \quad \text{.....(4-5)}$$

$$\beta = \tan^{-1}[A / (B+1)] \quad \text{.....(4-6)}$$

where P is the profile of maximum decentering distortion along the axis defined by the anti-clockwise angle θ from the x-axis, and λ_y is the scale factor for the y-axis relative to a unit scale along the x-axis, β is the angular deviation from the orthogonality between x- and y-axes.

The object space control coordinates are incorporated into the solution separately in X, Y and Z, therefore they may be either horizontal or vertical in the allowing condition equation:

$$G_{Xi} \equiv X_{Gi} - X_i = 0$$
$$G_{Yi} \equiv Y_{Gi} - Y_i = 0 \qquad \text{.....(4-7)}$$
$$G_{Zi} \equiv Z_{Gi} - Z_i = 0$$

where $(X_G, Y_G, Z_G)_i$ are the known coordinates of control point i.

Consider the condition equations as a function of two sets of unknowns $\bar{x}_1, \bar{x}_2$ and the observed quantities L:

$$\bar{X}_1 = (\Delta x, \Delta y, x_0, y_0, c, X_c, Y_c, Z_c, \omega, \phi, \kappa)$$
$$\bar{X}_2 = (X, Y, Z) \qquad \text{.....(4-8)}$$
$$L = (x, y, X_G, Y_G, Z_G)$$

then, equations (4-1) and (4-2) can be expressed in general form as:

$$F(\bar{X}_1, \bar{X}_2, L) = 0$$
$$G(\bar{X}_2, L) = 0 \qquad \text{.....(4-9)}$$

The system of condition equations(4-9) is non-linear, and any redundancy in observations lends itself to a least-squares adjustment. In order to avoid ill-conditioned normal equations the set of parameters under consideration is segmented into several uncorrelated subsets(sequential segmentation procedure). The solution of the normal equations is obtained by triangular factorization and back substitution.

4-2. Direct Linear Transformation(DLT)

4-2-1. Basic Concept and Solution Approaches

The innovation in the DLT approach is the concept of direct linear 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, Marzan & Karara, 1975]. As such, the DLT solution makes no use of fiducial marks. The method is based on the following pair of equations:

$$x_{ij} + \Delta x_{ij} = \frac{L_1 X_i + L_2 Y_i + L_3 Z_i + L_4}{L_9 X_i + L_{10} Y_i + L_{11} Z_i + 1}$$

$$y_{ij} + \Delta y_{ij} = \frac{L_5 X_i + L_6 Y_i + L_7 Z_i + L_8}{L_9 X_i + L_{10} Y_i + L_{11} Z_i + 1} \quad \text{.....(4-10)}$$

where

$$\Delta x_{ij} = dr_{xj} + dp_{xj}$$

$$\Delta y_{ij} = dr_{yj} + dp_{yj} \quad \text{.....(4-11)}$$

$(L_1 \ldots L_{11})_j$: unknown transformation parameters for photograph j

Besides systematic errors, the observed comparator coordinates will contain random errors v_x and v_y, respectively, and equations (4-10) for point i in photograph j should be rewritten as

$$(x + v_x + \Delta_x)(L_9 X + L_{10} Y + L_{11} Z + 1) - (L_1 X + L_2 Y + L_3 Z + L_4) = 0$$

$$(y + v_y + \Delta_y)(L_9 X + L_{10} Y + L_{11} Z + 1) - (L_5 X + L_6 Y + L_7 Z + L_8) = 0 \quad \text{.....(4-12)}$$

Two different approaches can be used in the formulation of a least-squares solution by means of equation(4-12). Each object point which has known object space coordinates(X,Y,Z) gives rise to two observation equations.

If we take into account only K_1, 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 (4-10) is computed by an iterative least squares adjustment. The number of unknowns carried in the solution will depend on how much systematic errors are corrected for in the solution.

After the unknown transformation parameters are determined, the object space coordinates of targets are calculated. The following mathematical derivation is summarized from the article by Marzan & Karara(1975). The systematic error($\Delta x, \Delta y$) are computed from equation (4-11) and applied to the observed coordinates x, y. If we put

$$\bar{x} = x + \Delta x$$
$$\bar{y} = y + \Delta y, \qquad \text{.....(4-13)}$$

then, from equation (4-12), we have the relationships:

$$(\bar{x}L_9 - L_1)X + (\bar{x}L_{10} - L_2)Y + (\bar{x}_{11} - L_3)Z + (\bar{x} - L_4) = 0$$
$$(\bar{y}L_9 - L_5)X + (\bar{y}L_{10} - L_6)Y + (\bar{y}_{11} - L_7)Z + (\bar{y} - L_8) = 0 \qquad \text{.....(4-14)}$$

In each photograph, therefore, we can write for each point on a set of equations (4-14). If there are p photographs used in the solution, we will have 2p equations to compute the unknowns X, Y, Z, the object space coordinates of a point . The number of degrees of freedom will be r=2p-3.

In matrix notation, the pair of condition equations we can write for point i in photo j are

$$V_j + B_j\Delta_i + C_j = 0 \qquad \text{.....(4-15)}$$

where

$$V_j = \begin{bmatrix} V_x \\ V_y \end{bmatrix}, \qquad \Delta_i = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ii}$$

$$B_j = \begin{bmatrix} (xL_9 - L_1)(xL_{10} - L_2)(xL_{11} - L_3) \\ (yL_9 - L_5)(yL_{10} - L_6)(yL_{11} - L_7) \end{bmatrix}_j$$

$$C_j = \begin{bmatrix} x + A\Delta x - L_4 \\ y + A\Delta y - l_8 \end{bmatrix}$$

$$A = L_9 X + L_{10} Y + L_{11} Z + 1$$

For p photographs, we have

$$V + B\Delta_i + C = 0 \qquad \text{.....(4-16)}$$

A least squares solution will again give

$$\Delta_i = -(B^T W B)^{-1} B^T W C \qquad \text{.....(4-17)}$$

In equation (4-17), W, the weight associated with each condition equation, is obtained by the law of propagation of variances from equation (4-12) [see Marzan & Karara, 1975].

The statistical errors in the above solution are usually assessed in terms of the variance-covariance matrix of the computed coordinates (or unknown parameters) given by

$$m_{\Delta i} = \overline{\sigma}_0^2 \left[B^T W B \right]^{-1} \qquad \text{.....(4-18)}$$

$$\overline{\sigma}_0^2 = \left[\frac{V^T W V}{r} \right]$$

4-2-2. Advanced Approaches Related to DLT

(1) The 11 Parameter Solution

In the DLT solution, the 11 transformation parameters L_i are considered independent. The 11 parameters L_1 through L_{11} can be physically interpreted in terms of the photograph's interior and exterior orientations. The fact that only 9 of these DLT parameters are independent is not accounted for in the approach indicated by equation (4-10). For example, L_4 can be expressed as a linear combination of L_1, L_2 and L_3, and L_8 is linearly dependent on L_5, L_6 and L_7. However, this approximation is rectified in so called 11-parameter solution through the use of nonlinear parameter constraints [Bopp & Krauss, 1977]. These constraints are

$$(L_1^2 + L_2^2 + L_3^2) - (L_5^2 + L_6^2 + L_7^2) + (C^2 - B^2)/D = 0$$

$$(L_1L_5 + L_2L_6 + L_3L_7) - BC/D = 0 \quad \text{.....(4-19)}$$

where

$$B = L_1L_9 + L_2L_{10} + L_3L_{11}, \quad C = L_5L_9 + L_6L_{10} + L_7L_{11}$$

$$D = L_9^2 + L_{10}^2 + L_{11}^2$$

These constraints are incorporated in the solution as additional observation equations with zero variance. In other words, Bopp & Krauss (1977) tried to overcome the imperfection of DLT with the introduction of additional constraints on the parameters. Similarly, Dermanis(1994) derived the photogrammetric inner constraints, which are used for the derivation of free network solutions in close-range applications.

(2) An Iterative Linear Transformation Algorithm

In order to improve the performance of the conventional DLT program, Naftel and Boot(1991) proposed the following modifications:

a. An initial solution to the unknown object space coordinates through equations (4-10) and (4-17), assuming unit weight , can be readily obtained using one of the standard mathematical subroutines.
b. The initial solution for the unknown object space coordinates is then fed back into Marzan and Karara's program, along with the known control point coordinates which are maintained as constant. Thus, photo-coordinate observations to non-control points are used in the adjustment leading to the solution of the calibration parameters, which has a great degree of redundancy.
c. The compound DLT solution serves as a better trial approximation to the coordinates of the unknown points and allows the procedure to be repeated in an iterative fashion.

Naftel & Boot(1991) showed that the iterative linear transformation procedure provides considerably reduced RMS errors in comparison to the original DLT results.

4-3. Evaluation of the Results

In order to evaluate the suitability or effectiveness of a particular adjustment technique, some criteria for the results must be established. It is generally agreed that quality has three aspects: accuracy, precision and reliability. These concepts can overlap in some situations, but in general each measures a separate, independent aspect of the adjustment and each should be

calculated and examined when a photogrammetric solution is designed and executed. Each factor is affected by the geometry and the execution of a photogrammetric task.

4-3-1. Accuracy

The accuracy of measurements is defined by the degree of closeness to the true value. Accuracy is influenced, not only by a random error component of the measurement, but also by the bias created by uncorrected systematic errors. Where an object space control field has been surveyed by conventional means, check point coordinates can be used to express photogrammetric adjustment accuracy in terms of the root mean squares errors μ of the object point coordinates. The mean square error μ^2 is defined as following the expectation:

$$\mu_X^2 = E\left[(X-\tau)^2\right] \quad \text{.....(4-20)}$$

where X represents the object point coordinates as adjusted, and τ is its true value surveyed by conventional means.

4-3-2. Precision

Precision describes the statistical quality of the estimated parameters, if the a-priori assumptions(functional and stochastic relations) of the adjustment model are considered to be true. Hence the variance-covariance matrix contains all the information concerning the precision of the parameters obtained. The variance-covariance matrix may be transformed to give more useful numbers by calculating a standard deviation-correlation matrix from it. The diagonal elements are then the standard deviations of the parameters, calculated as the square root of the variances, while the off-diagonal terms are the correlation coefficients between the parameters.

The following very popular precision measures(global indicators) are obtained by using the traces of the corresponding variance-covariance matrices [Grün, 1978] :

$$\overline{\sigma}_X^2 = \frac{tr(\Sigma^X)}{n_X}$$

$$\overline{\sigma}_Y^2 = \frac{tr(\Sigma^Y)}{n_Y} \qquad \text{.....(4-21)}$$

$$\overline{\sigma}_Z^2 = \frac{tr(\Sigma^Z)}{n_Z}$$

Σ : variance-covariance matrix of parameters

$\Sigma^X, \Sigma^Y, \Sigma^Z$: corresponding parts of Σ for X,Y,Z

n_X, n_Y, n_Z : numbers of X,Y,Z coordinates.

Near-homogeneous precision is often desired for the X,Y and Z coordinates, and a single estimator σ_C^2 can be employed to express the mean precision of the object coordinates :

$$\overline{\sigma}_C^2 = \frac{tr(\Sigma)}{n} \qquad \text{.....(4-22)}$$

where n is the number of the object point coordinates.

It should be noted here that the precision measures given here are with respect to a given coordinate datum. The coordinates obtained will in general be correlated, that is, the covariance matrix for each point will have off-diagonal terms. The rotation of the coordinate system would therefore result in changes in the diagonal elements of the variance-covariance matrix, which are the variances of the point coordinates. One approach to avoid this situation is to use the so-called free-net adjustment in which the influence of the datum parameters is minimized [McGlone, 1989, Dermanis, 1994].

4-3-3. Reliability

In addition to specifying adjustment quality in terms of precision, the concept of reliability was employed by Baarda(1967). Reliability defines the quality of the adjustment model with respect to the detection of model errors. Reliability is concerned not only with the random errors, as is precision, but also with bias which can imply both gross errors and systematic errors. Currently , the term reliability refers mainly to blunder detection. A popular method of checking for gross errors after an initial adjustment is to compare the weighted residuals with $\bar{\sigma}_0^2$, the reference variance, in equation (4-18). For example, if

$$\left|v_i w_{ii}^2\right| > 3\bar{\sigma}_0 \quad \text{.....(4-23)}$$

we suspect a gross error in the i th observation. However, such a technique assumes that all the residuals can be represented by a common a-posteriori variance.

Grün(1978) has adopted Baarda's data snooping technique for gross error(blunder) detection in photogrammetry. The reliability of a photogrammetric network adjustment is assessed in terms of internal and external reliability. The internal reliability gives the magnitude of a blunder in an observation (∇l_i) which is just non-detectable on a certain probability level. In the following equation it is assumed that only one blunder appears in the network:

$$\nabla l_i = \sigma_0 \frac{\delta}{w_i \sqrt{q_{v_i v_i}}} \quad \text{.....(4-24)}$$

δ : non-centrality parameter of the data-snooping test

w_i : weight of observation l_i

$q_{v_i v_i}$: i th diagonal element of cofactor matrix Q_{vv}

Furthermore, Baarda(1967) developed a criterion for testing the residuals of the adjustment(data snooping).

$$u_i = \frac{-v_i}{\sigma_0 \sqrt{q_{v_i v_i}}} \qquad \text{.....(4-25)}$$

The acceptance interval for u_i is

$$-F^{1/2}(1-\alpha_0,1,\infty) < u_i < F^{1/2}(1-\alpha_0,1,\infty) \qquad \text{.....(4-26)}$$

α_0 : type I error size for test criterion u_i

In contrast to the simple criterion, of equation (4-23), the design matrix B is taken into account in data snooping(equation (4-25)), thus provides a much more satisfactory criterion for the detection of gross errors.

The external reliability indicates the effect of a non-detectable blunder on the estimated quantities (∇x_j).

$$\nabla x_j = (B^T W B)^{-1} B^T W \nabla l_i \qquad \text{.....(4-27)}$$

B : design matrix of the model

For bundle adjustment, the internal reliability can be considered to be defined in the image space and the external reliability in the object space.

5. DYNAMIC MONITORING

The processing of an image sequence involving motion has become increasingly important. The following is a partial list of applications:

1) Military problems - Target detection and recognition.
2) Industrial problems - Dynamic monitoring of industrial processes. Robot vision.
3) Commercial problems - Bandwidth comparison of picture phone video signals.
4) Medical problems - Study of cell motion by microcinematography.
5) Meteorology - Cloud tracking.
6) Transportation - Highway traffic monitoring.

In this report, dynamic monitoring means the measurement of 3-dimensional coordinates of objects which are moving with time, based on sequential video images. The trajectories of moving objects are obtained by repeating measurements of 3-dimensional coordinates of the objects sequentially with a specific time interval. Therefore, sequential images are a prerequisite. Since the 1970s, solid-state cameras, instead of high speed film cameras with magazine or movie camera, are used to take sequential images. This offers new possibilities for the image formation process, since the images containing the feature of interest can be acquired in a sequential manner at about thirty images per second. In conventional photogrammetry, 3-dimensional coordinates of objects, staying stationary, are obtained from stereo images. A stereo image is formed from two consecutive images of an object photographed at two exposure stations separated by a certain distance. However, it is impossible to photograph moving objects at two station with one camera at the same time. Various methods to overcome this obstacle were devised.

One of them is deriving 3-dimenional coordinates of objects just from sequential single images, instead of stereo images. The other one is taking the left and right scene in a image simultaneously with a camera. The last one is

taking the left and right image using two(or more) cameras separated by some base length(s) at the same time. Because just one camera is used to take a picture in the former two approaches, a pair of images (or a single image) are captured simultaneously. However, in the last approach, it is necessary to control the exposure of two(or more) cameras, starting and ending at exactly the same time. The former two approaches are called synchronization-free approaches in this report.

Here a mathematical solution for deriving 3-dimentional coordinates from sequential single images is introduced. On the other hand, since the mathematical principle of deriving 3-dimentional coordinates from stereo images is same as in conventional photogrammetry, only the methods of capturing the left and right scenes on a single image with a camera are reviewed. Though the above synchronization free approaches with a single camera provide 3-dimentional coordinates, their range of use is limited for its lower accuracy, its narrow field of view and so on. As it were, stereo images from a multi-cameras system are more desirable only if the proffer synchronization is provided. Therefore internal and external synchronizing method are discussed. Also, for tracking moving objects, the strategy of searching for the same points in the sequential images is briefly described.

5-1. Synchronization-Free Approaches

5-1-1. Three-Dimensional Coordinates from Sequential Single Images

We shall assume in this sub-chapter that we are working with two time-sequential images(frames) to find the amounts of motion that the objects have undergone from time t_1 to time t_2. The geometry of the problem is sketched in Fig. 5-1. The object-space coordinates are denoted by (X, Y, Z) and the image space coordinates denoted by (x, y). Huang & Tsai(1981) suggested a two step method for estimating the motion parameters. First, the image-space shifts of a

number of points on the same rigid body are estimated. Then, the motion parameters of the rigid body are determined from these image-space shifts.

1) Estimating Image Space Shifts

The image space shift can be estimated in a variety of ways(by a human operator, or by cross-correlation, or by the differential method). When the differential method is used, we have the following approximate differential equation at each image point (x, y).

$$\Delta f = -\frac{\partial f}{\partial x}\Delta x - \frac{\partial f}{\partial y}\Delta y \qquad \text{..... (5-1)}$$

where $f_1(x,y)$: Image intensity at (x, y) and time t_1

$f_2(x,y)$: Image intensity at (x, y) and time t_2

$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right)$: spatial gradient of $f_2(x,y)$ at (x, y)

Δf : frame difference at(x, y) = $f_2(x,y) - f_1(x,y)$

If a small number of nearby points can be assumed to have the same shifts $(\Delta x, \Delta y)$, then by calculating Δf and $(\partial f / \partial x, \partial f / \partial y)$ at each point, we can get a number of equations (5-1). By solving them using least squares, we get an estimation of $(\Delta x, \Delta y)$.

2) Determining Motion Parameters

Any three-dimensional motion of a rigid body can be decomposed into rotation around the axis passing through the origin, and translation. Therefore, we have the following basic equation.

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \qquad \text{..... (5-2)}$$

where X,Y,Z : object coordinates of point P at time t_1

X',Y',Z' : object coordinates of point P at time t_2

$\Delta X,\Delta Y,\Delta Z$: translation of a object in the object space from time t_1 to t_2

R : rotation matrix

Let n_2,n_2 *and* n_3 be the direction cosines around the axis of rotation, and ϑ the amount of rotation between the two frames, then the rotation matrix becomes:

$$R=\begin{bmatrix} n_1^2+(1-n_1^2)\cos\vartheta & n_1n_2(1-\cos\vartheta)-n_3\sin\vartheta & n_1n_3(1-\cos\vartheta)+n_2\sin\vartheta \\ n_1n_2(1-\cos\vartheta)+n_3\sin\vartheta & n_2^2+(1-n_2^2\cos\vartheta & n_2n_3(1-\cos\vartheta)-n_1\sin\vartheta \\ n_1n_3(1-\cos\vartheta)-n_2\sin\vartheta & n_2n_3(1-\cos\vartheta)+n_1\sin\vartheta & n_3^2+(1-n_3^2\cos\vartheta \end{bmatrix}$$

..... (5-3)

Considering the characteristic of the direction cosines $(n_1^2+n_2^2+n_3^2=1)$, and assuming that the amount of rotation , ϑ, is small, then

$$R=\begin{bmatrix} 1 & -n_3\vartheta & n_2\vartheta \\ n_3\vartheta & 1 & -n_{1\vartheta} \\ -n_2\vartheta & n_1\vartheta & 1 \end{bmatrix}. \quad \text{..... (5-4)}$$

Substitute equation (5-4) into equation (5-2), and after some algebraic manipulation, we have

$$\Delta x=\frac{-(fn_3y-f^2n_2-n_2x^2+n_1xy)\vartheta\overline{Z}+f\Delta X'-fx}{-(n_2x-n_1y)\vartheta\overline{Z}+f\overline{Z}+f}$$

$$\Delta y=\frac{-(fn_3x-f^{2x}n_1+n_1y^2-n_2xy)\vartheta\overline{Z}+f\Delta Y'-fy}{-(n_2x-n_1y)\vartheta\overline{Z}+f\overline{Z}+f} \quad \text{..... (5-5)}$$

where f : camera focal length

$$\Delta Z' \cong \Delta Z / f$$

$$\Delta X' \cong \Delta X / \Delta Z'$$

$$\Delta Y' \cong \Delta Y / \Delta Z'$$

$$\overline{Z} = Z / \Delta Z .$$

If we pick an image-space point and estimate the shift, then equation (5-5) gives us 2 equations with 6 unknowns, $n_1, n_2, n_3, \Delta X', \Delta Y'$ *and* $\overline{Z}$. Unfortunately, the equations are non-linear. Each new image-space point gives 2 new equations and one new unknown ($\overline{Z}$). Therefore, we need a minimum of 5 image-space points. In that case, we have 10 non-linear equations with 10 unknowns. We can, of course, take more points and find the least-squares solution.

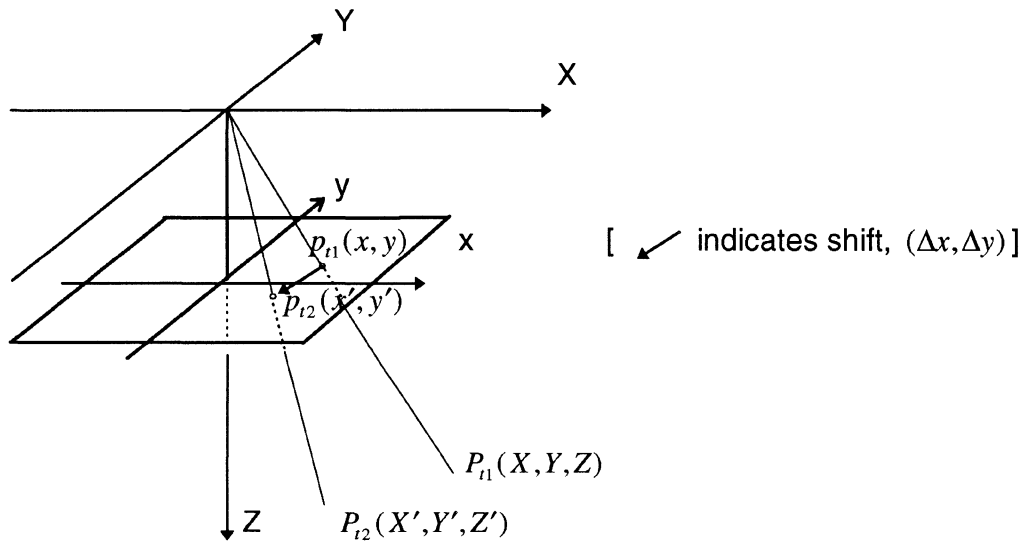


Fig. 5-1. Basic Geometry for Three-Dimensional Motion Estimation.

5-1-2. Single Camera Systems for Sequential Stereo Images

A single camera system, as the name implies, requires a single exposure for the three-dimensional reconstruction of the object photographed. Such procedure does not preclude the use of photogrammetry's strongest implement, namely that of the stereoscopy. In conventional photogrammetry the two exposure are separated by the base, from which the object is being photographed. If it were possible to transfer this stereoscopic base into an object space, to replace the the object at one end, and the identical replica of it at the other end of the base, then the single exposure station would suffice to produce a photographic record which can be investigated three-dimensionally. The use of the system is limited to close-range applications involving reasonably short base lengths. Here we have two methods, mirror photogrammetry and beam splitter photogrammetry.

1) Mirror Photogrammetry

The practical realization of mirror photogrammetry is achieved by placing a reflecting surface(e.g. mirror) in close proximity to the object itself. Fig. 5-2 is a presentation in plan of the relationship between object space and camera station. Such an arrangement enables us to obtain simultaneous exposures of an instant on one photograph. The mirror is set up to one side of the exposure axis. By the law of the reflection this mirror surface produces a replica of the object. The apparent object is recorded by the camera as a left-right reversed image. The apparent second camera station is located at a distance equal to twice the perpendicular distance of the original camera station from the mirror surface. This distance represents an artificial base length. The two images recorded of the real and as the apparent object are, therefore, two different central perspectives, and can be subjected to stereoscopic examination[Gruner,

1955]. The base length increased twice as fast as the displacement of the mirror surface from the axis of the camera.

As seen in Fig 5-2, when the exposure station axis and the mirror surface are parallel, this establishes what is known as normal case photos. By rotating the mirror surface, we can get convergent photos. Faig(1972) stated that mirror photogrammetry is suitable for most close-range objects, including those in motion or being unstable.

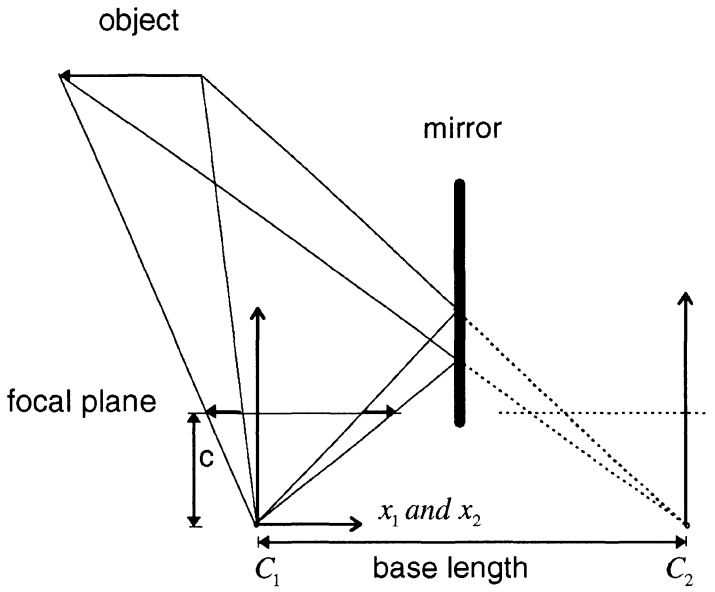


Fig. 5-2. Geometric Relationships of Mirror Photogrammetry

2) Beam Splitter Photogrammetry

Another system which can be used to obtain three-dimensional information from a single photographic exposure is the beam splitter attachment, shown in Fig. 5-3. A beam splitter consists of two surface-coated plane mirrors and a surface-coated 90º prism. When it is mounted in front of the camera lens, two separate bundles of rays are formed so that a photographed object is

recorded in two different locations on the photograph. This photograph contains, therefore, the same information as offered by two photographs taken by simultaneously operated, perfectly synchronized cameras, which is of particular interest in photogrammetric studies of high speed processes. The base length of the stereo photographs, obtained by this beam splitter attachment, is approximately the distance between the outside mirrors.

This system was used in the photogrammetric determination of particle velocity vectors in a vertical two-phase solid-gas flow[van Wijk & Ziemann, 1976]. Lu Jian et al.(1992) attached a beam splitter in front of CCD camera and used the system for tracking human gait.

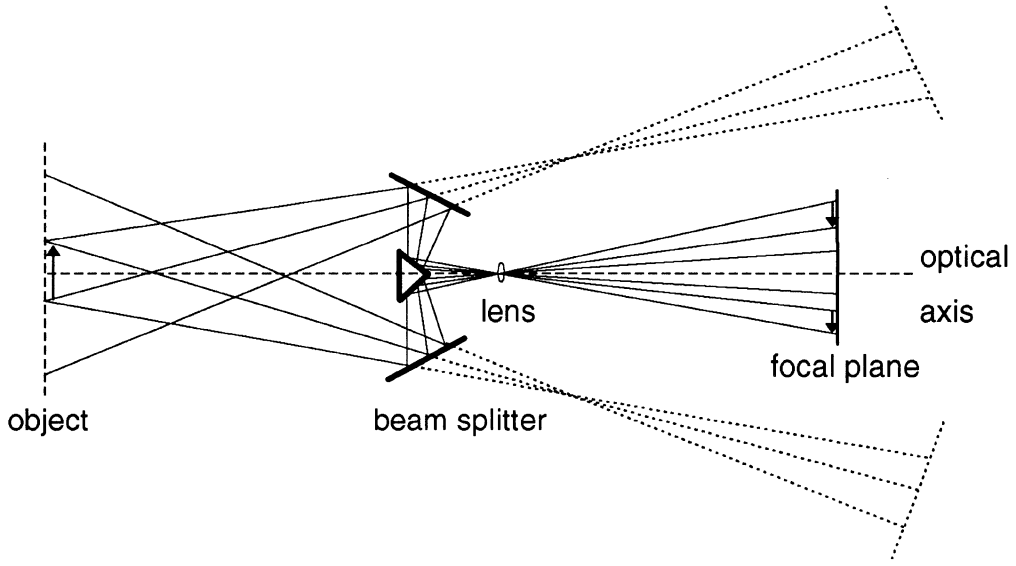


Fig. 5-3. Geometric Relationships of Beam Splitter Photogrammetry

5-2. Synchronization for Multi-Cameras Systems

When we photograph moving objects with two(or multi) cameras, it is very important to synchronize the left and right images. Otherwise, each image

contains the scene of the objects in different locations, since the objects are moving. It eventually causes 3-dimensional coordinate errors. In order to diminish their components to lower than allowable values, the cameras must be synchronized within a certain limit. When CCD cameras are used as sensors, synchronization between the cameras is achieved by an internal electric signal or by an external device flashing visible light. Let us call it a-priori synchronization because the sequential images are carrying the information necessary for synchronizing between photographs.

The principle of internal synchronization methods is that a master camera sends its pixel clock signal to the other(slave) cameras to control exposure of all cameras, or the host computer equipped with an image processing board, sends its internal clock signal to the all cameras connected in order to synchronize the successive exposures of the cameras. On the other hand, Baltsavias & Stallmann(1991) attached an electronic device between the cameras and the VCR. This device periodically sends a signal to add a vertical line of high contrast and a binary pattern to the video signal. The binary pattern provides a code for the sequence of each image during recording and permits safe identification of each image during the digitization.

Another method of synchronization is stroboscopic. Properly speaking, a stroboscope is an instrument for observing the successive phase of a periodic motion by means of a light periodically that is interrupted. The early stroboscope s used mechanical modulation of the light, but with the development of gas discharge tubes, the successive flashes of light were achieved electrically. In stroboscopic synchronization the cameras are considered to be shutterless, and the extremely brief duration of the flash arrests the motion of the object.

In addition to the above photographic methods, every sign displayed periodically could be used for synchronization. For example, a stopwatch refreshes its digits every 1/100s. If the digits can be differentiated on the image, it can be used as a code of the sequence.

Even though the sequential images were captured without a-priori synchronization, we can afterwards deduce the synchronized images by using a video recording/playing system. Actually, the frame advance function of the VCR could also be used for that purpose, provided that a significant starting point was identified in each image.

5-3. Object Tracking

A human operator could easily identify certain objects(or targets) in each sequential image. The requirement for very fast data acquisition and processing is generally dictated by the facts that certain measurement problems could not be solved otherwise and/or that a replacement of a human operator by automatic system turns out to be more precise and reliable, giving faster and less expensive results. Typical applications in industry are found for instance when dealing with moving objects, in quality control and inspection of products without stopping the production line, in manufacturing, in transportation and navigation, and in surveillance and object identification. Matching algorithms as described in chapter 3-2 is also used for tracking targets in sequential images.

From the operational point of view, automatic tracking has some advantages over other matching for relative orientation or target location. Based on the assumption that image motion is limited to a few pixels from one image to the next, the position of the target in the current image is used as an approximation for its position in the next one. Therefore a small search area centered on the last position of the target is searched linearly. On the other hand, since the target is moving, the gray values of the target are subject to change in each sequence, and occlusions are inevitable.

6. EXPERIMENTS

6-1. Preliminary Test for the Video System

When using a non-metric camera for measurements, usually the vendor does not provide the necessary specifications required for data reduction and processing. For that reason, the basic characteristics of the camcorder and the quality of the video image had to be known, before three-dimensional measurements were undertaken. First, it was necessary to confirm whether the video measuring system was structured correctly and works as expected. Secondly, we needed data on the radiometric characteristics and the lens distortion of the camcorder. Thirdly, we are interested in qualify of imaging of the moving objects by the camcorder. In order to get these objectives, we designed a simple test, and conducted the experiments described below.

6-1-1. System Set-Up

Basically I set up our 3 dimensional measuring system with hand-held amateur camcorders, VCRs, and a frame grabber used in the PC. They all have a common drawback of low metric characteristics. On the other hand, it is easy to access and manipulate these tools, their prices are low, and the renewal interval is shorter than for metric equipment.

In order to get and record stereo images, one S-VHS and one VHS camcorder were connected to two S-VHS VCRs. Then a VCR was connected to a PC installed frame grabber, and sequential images were captured manually. As an image processor, IDRISI for Windows was used. Though this software was originally developed for GIS, IDRISI has a lot of modules to process digital images on a PC. After getting image coordinates of targets in each pair of sequential images, their three dimensional ground coordinates were acquired based on DLT on the IBM main frame. Table 1. lists the hardware and software

of the system used to obtain the images of moving objects and of the target locations.

Table 6-1. System Components for Video Image.

Hard/Software	**Maker, Model**	**Specifications**
Video Camera I	Panasonic, AG-455P	1/3" CCD, S-VHS, S-Video, ½" tape
II	Samsung, H-33	1/3" CCD, VHS, S-Video, Hi-8 Tape
VCR	Panasonic, AG-1970	S-VHS, S-Video
Frame Grabber	CreativeLab, VideoBlaster SE	Resolution: 640*480(VGA)
Video Tape	Sony, Sony V	S-VHS
PC	Dell, Dimension P100C	Intel 100MHz Pentium, 16Mb Ram
Image Processor	Clark University, IDRISI	Window 95, Compatible with TIFF

6-1-2. Procedure for Determining Sub-Pixel Target Locations.

After recording an image on the S-VHS VTR, its analogue signal was digitized by the frame grabber. Up till now, most of the frame grabbers used in a PC's cannot process images with the NTSC standard rate. Moreover the frame grabber is not usually compatible with Window 95. Therefore it was inevitable to obtain sequential images with Window 3.1 in an off-line mode. Listed below is the detailed procedure and its specifications to obtain target locations from video images.

(1) Frame Grabbing
Frame Grabber : Video Blaster SE
Input Analogue Image : NTSC(RS170) TV signal

Output Digital Image Format : BIP(24bit Band-Interlaced-by-Pixel)
Resolution : 788(H)*468(V)(on Super VGA 800*600)
628(H)*372(V)(on VGA 640*480)
Image Type : B/W(or RGB)
Save File Format :Tagged Image File(TIF)

(2) Conversion from BIP to BSQ format : BIPIDRIS (in IDRISI for Windows)

	Super VGA(800*600)	VGA(640*480)
BIP Total File Size(Bytes)	1,106,352	703,688
Header Size(Bytes)	4532	2840
Bits per Pixel per Band	1718	1718
BSQ Total File Size(Bytes)	368,784	233,616
Number of Band	3(1:R, 2:G, 3:B)	3(1:R, 2:G, 3:B)
Gray Level	0(Black) ~ 255(White)	0(Black) ~ 255(White)

(3) Noise Removal : Median Filter

(4) Determine Thresholding Limit : Histogram

(5) Reverse Gray Value : Reclass

(6) Target Segmentation(Thresholding) : Overlay

(7) Feature Extraction : Window

(8) Target Location : Center

6-1-3. Basic Characteristics of the Video System

The radiometric and geometric properties of all elements involved in image acquisition and information extraction must be investigated in order to locate sources of degradation. As Curry et al(1986) have noted, radiometric criteria include sensor response linearity, system noise, and system stability over time. In these experiments, the dark signal non-uniformity(noise) and photo response non-uniformity were basically used to assess the radiometric performance. System noise is defined as fluctuations in pixel gray levels caused by random and systematic perturbations in the CCD sensor and A/D converter.

In order to obtain radiometric characteristics of camcorders, VCR, and frame grabber, video images were taken under various conditions.

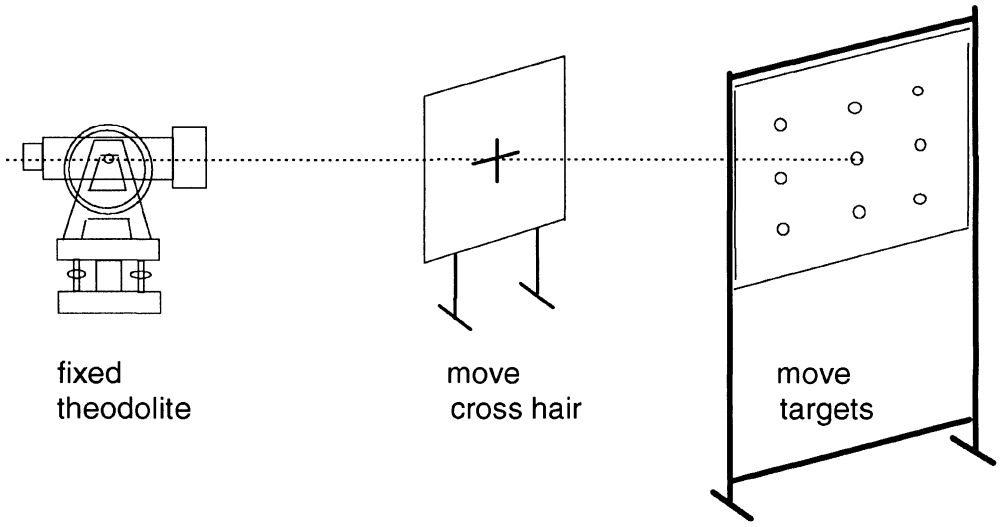
At first, a pure dark image of the frame grabber was acquired without providing a video signal from VTR. Dark images of the camcorders were acquired by running the camcorders in the dark. Also white images were acquired by photographing white paper under bright illumination.

As Bayer(1992) pointed out, the temporal stability, the spectral characteristics, and the distribution of the light intensity on the object affect the measurement accuracy in several ways. The most obvious problem associated with the illumination is the variation of its intensity due to inherent properties of the illumination and/or shadows. Only the gradient induced across a feature of interest is of importance(Global differences do not lead to positional changes with the typical target location algorithms like Least Squares Matching). In addition to the above phenomena, the shutter speed influences the image quality, too. So for black targets on white background picture were taken in the room and outdoors with different shutter speeds. That gives us some information about the interaction of illumination and shutter speed.

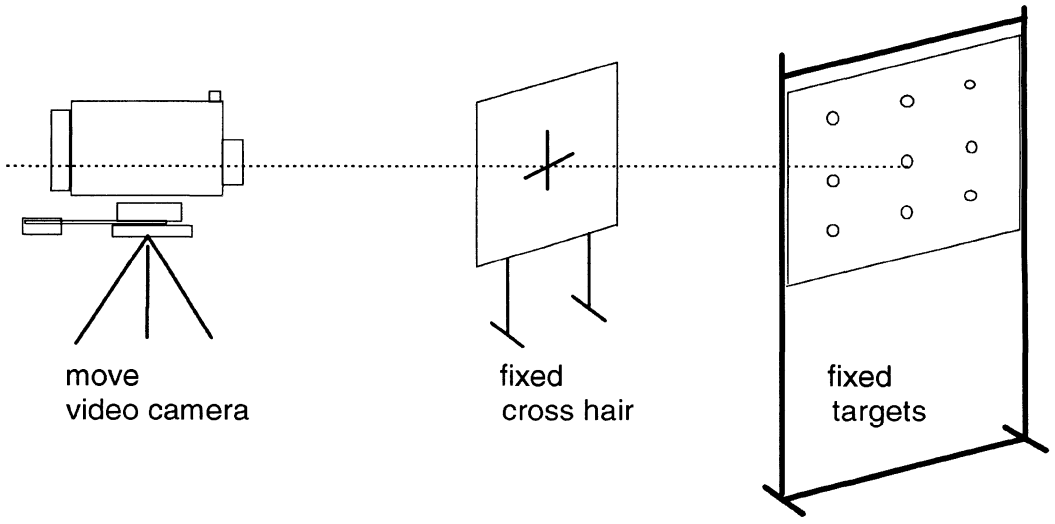
Lastly, a black and white(B/W) resolution test chart were captured under the same photographing condition with two camcorders. Even if we take images of the B/W objects, we can deduce a color band from them. Red, green, and blue band images were made by the following procedure. After taking a color image with the video camera, a digital image was captured in BIP format using the frame grabber. Then this BIP image was converted into BSQ format with 3 bands using the BIPIDRISI module in IDRISI for Windows. The B/W band can be obtained optionally in the course of frame grabbing.

There are many precise methods to find lens distortion. We opted for an on-the-job calibration, using a plane board with 5×5 circular targets as control points. To coincide the optical axis of the camera with the center target of the board, a cross hair drawn on a glass plate was placed in front of the target board. Using a theodolite, the central target and the cross hair on the glass plate

were aligned. Then the camcorder was adjusted such that its optical axis coincided with this line(Fig 6-1). As shown in Fig. 6-2, target board images were acquired at different object distances.



(A) Aligning a cross hair on a glass plate with a center target using a theodolite.



(B) Aligning a video camera with the optical axis.

6-1. Fixing the Optical Axis of a Camera Vertically to Target Board

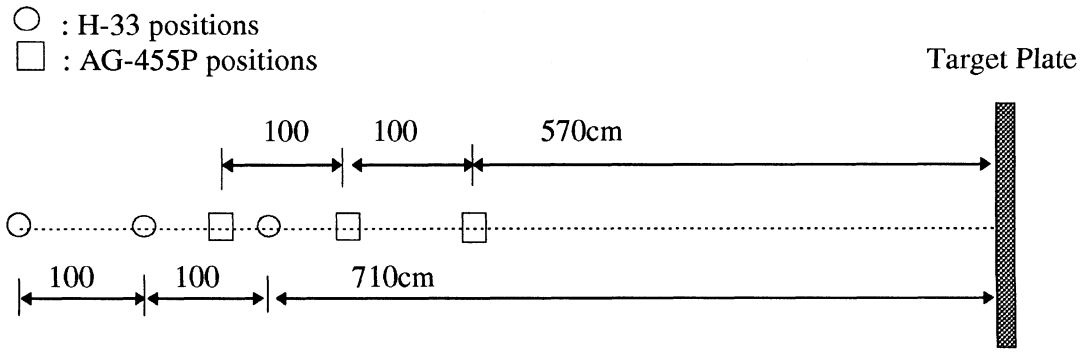


Fig. 6-2. Location of Cameras and Target Plate

6-1-4. Image of Moving Objects

As we know from aerial photography, the relative motion between the sensor and the object causes image blur. Blurred images degenerate the measuring accuracy. When it comes to imaging moving objects, image motion depends on the speed of the moving object, the pixel size in direction of the motion, the object distance, camera focal length, and shutter speed. Fine pixel sizes give us cleaner images of moving objects. The object distance and focal length relate to the image scale. Large scale images of moving objects give us more blurred images. The sensor pixel size is fixed once a camera was chosen, and usually the image scale is restricted due to the environments.

Shutter speed provides a tool to get a clean image of a moving object. To get practical recording of this phenomenon, I took images with specified shutter speeds, of the target board falling from different heights(Fig. 6-3). If the target board falls without friction in the stand, we can deduce the falling speed when it reaches the optical axis based on the law of gravity.

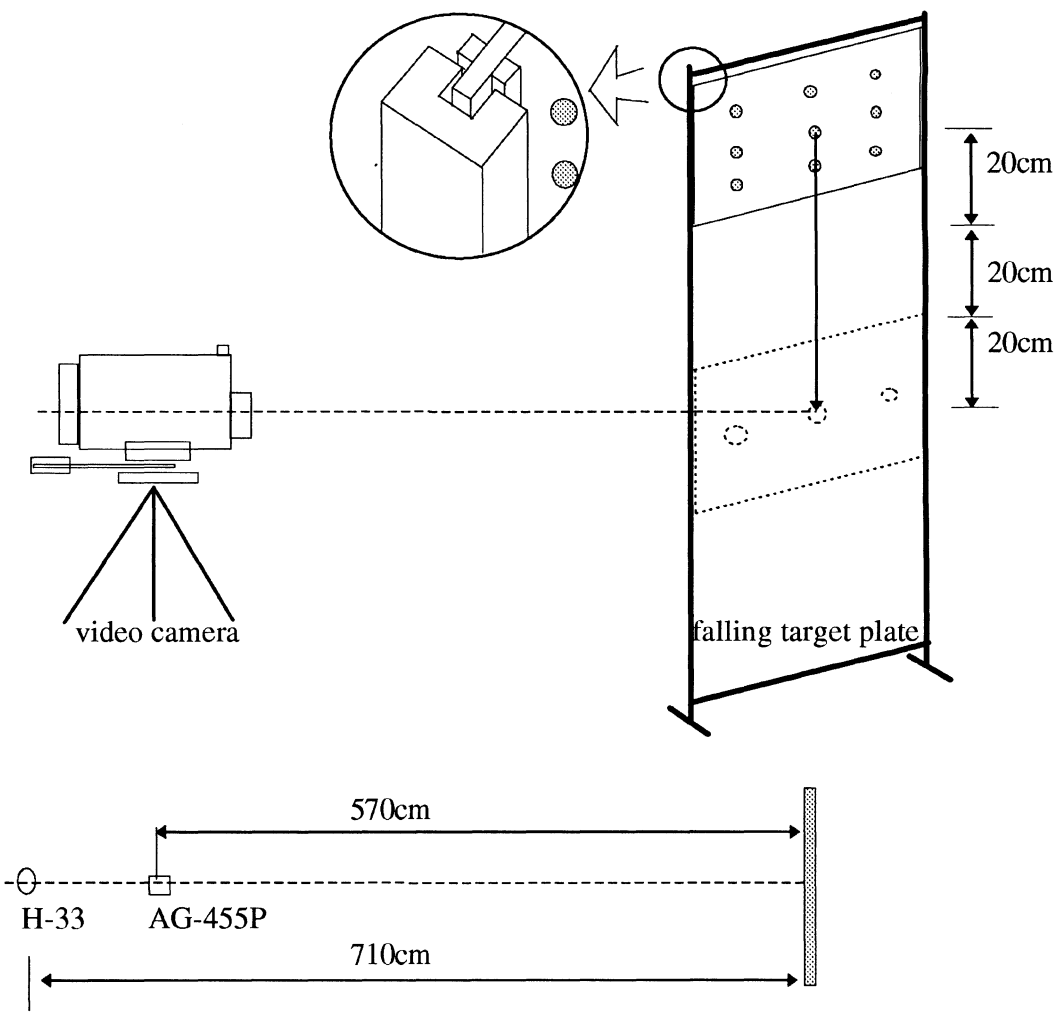


Fig. 6-3. Image Acquisition of a Falling Target Plate

6-2. 3-D Measurements of a Moving Car

6-2-1. 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 and vibrations. To achieve this objective, we selected a car as the vibrations and moving object, 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 ranging from small to large in the process.

As shown in Fig. 6-4, 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 the car. For control, I placed a 3 dimensional control frame in front of the car. The control frame has 20 control points distributed randomly, which were coordinated precisely with the Kern E2 Electronic Theodolite System [Faig et. al, 1996].

6-2-3. Imaging Geometry and Photographic Conditions

The adopted imaging geometry for a network is a 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 base to distance(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, I took images at three different base lengths, to compare the precision, and to chose the best base length for the test .

Based on the preliminary tests I fixed the shutter speed at 1/1000 s. Considering the field of the view of the camera(17~20°) and the size of the control frame sufficient to encompass the car, I decided on the camera stations. The diameter of the circular targets that could be imagined larger than 3 pixels in x, y directions was calculated. I used black circular stickers on white

white background as targets. The diameter of the stickers was 19mm and the width of the background was 3 times that of the stickers'. The average distance from the camera to the object was 2.7m. resulting in a photo scale of 1: 270. Additionally, to get a clean image of a car outdoors without artificial illumination, we needed sun shine. Unbalanced or poor illumination causes a lot of problems in target detection and location processes.

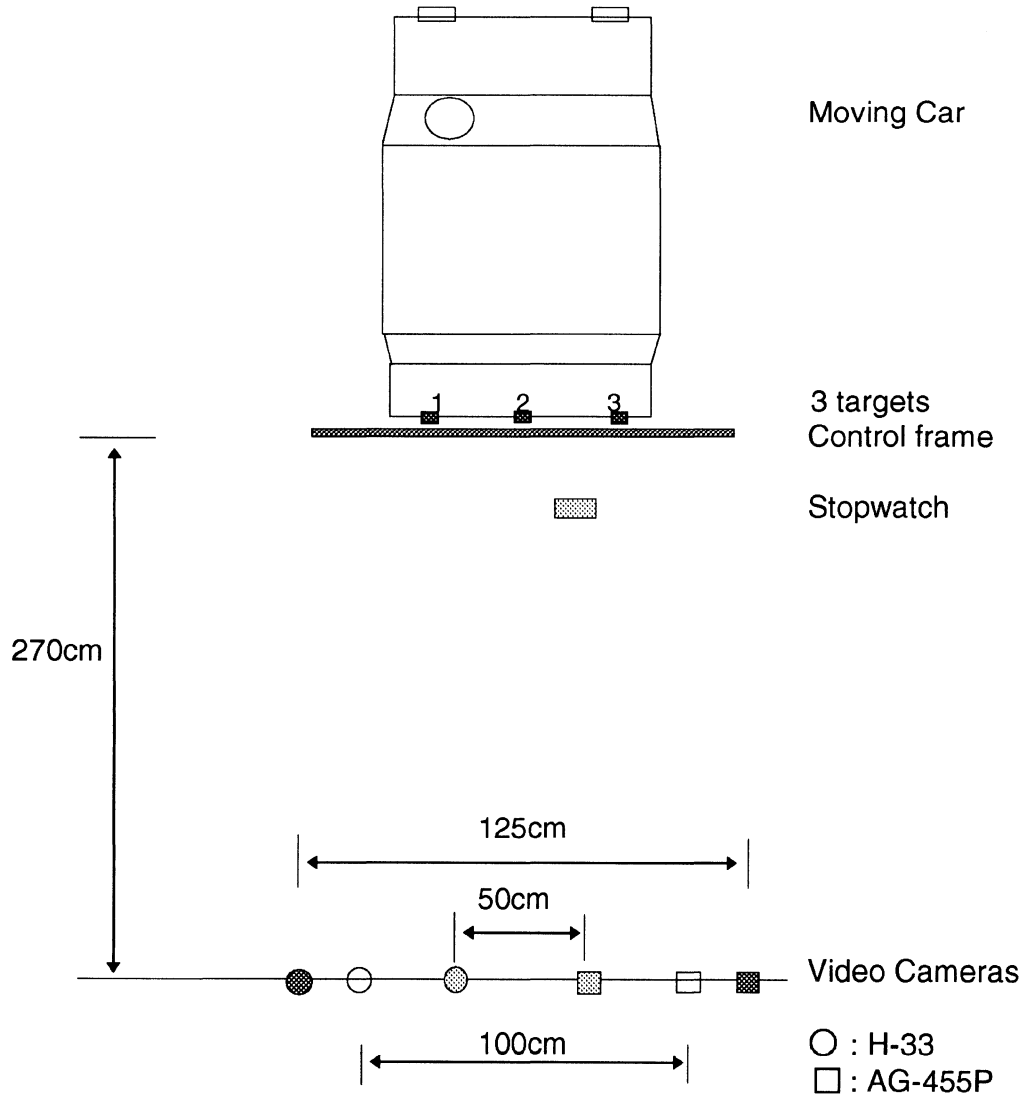


Fig. 6-4. Stereo-Image Acquisition of a Moving Car

6-2-3. Synchronization for Sequential Images

It is very important to synchronize the left and right images for stereo models of sequential images. Various synchronizing methods were described in section 5-2. In this research, a simple external synchronizing device, a stopwatch, was attached to the front of the control frame. It refreshes its digits every 1/100s, and its 1/100s digit was expected to be used as a code of the sequence. The size of 1/100s digit is 6mm* 3mm. When it was photographed at 1/270 scale, its size reduced to 22μm*11μm, which is roughly equivalent to 3*1.5 pixel. Therefore the digit in sequential images could not be differentiated as a code. In order to increase the scale of the digit, the stopwatch was put at a point between the video camera and control frame. However it was very hard to focus on both targets and stopwatch simultaneously.

As an a-posteriori method, there were no alternatives except the frame advance function of the VCR. Actually, Panasonic(AG-1970) S-VHS VCR provides a field(1/60s) advance function with a button. If a significant starting scene can be identified in the left and right images, it is possible to synchronize the remaining sequential images using the VCR. Therefore, while taking photographs the driver did several artificial abrupt motions in order to insert significant milestones into the sequential images. In addition to the driver's artificial motions, car door's opening and closing, and signal light turning on and off can be used for check points provided they are identified in the right and left images simultaneously.

7. TEST RESULTS AND ANALYSIS

7-1. Radiometric & Geometric Characteristics

7-1-1. Radiometric Characteristics

As it was specified in section 6-1-1, the main imaging components for these experiments are video cameras and a frame grabber. The analogue signals generated from the CCD sensor of a video camera flow into a frame grabber. The frame grabber converts these analogue signals into digital form. However, even though the camera does not provide any signal, the frame grabber may generate an intrinsic digital signal. Fig. 7-1 shows the intrinsic images for frame grabber. Its gray value range was between 6 and 14, and its distribution pattern was the same in all bands(Red, Green, Blue, and Black/White).

In order to detect the system noise of the cameras used, dark images were captured by running the camera in the dark(lens cap on) and then measuring the output for each pixel(Fig 7-2 and 7-3). Systematic noise is seen as a repeated spurious response by the sensor element. Dark images captured from both S-VHS and VHS video cameras look like that of the moon surface. Their gray value range in the B/W band was between 0 and 50, but S-VHS video camera shows a more concentrated histogram. When the gray value of a pixel in a dark image exceeded 3 times the standard deviation from the mean of the image, it was assumed to be a blemished pixel. The results are presented in Fig. 7-4 and Fig. 7-5. In these figures, black spots mean too low gray values, and white spots mean too high gray values. We found that every dark image of a video camera at an instance has some noise, but the position of noise pixels changed with time, and rarely fell on the same pixel position in different images.

The same experiments were conducted for a white image. White image means an image of white paper(Fig. 7-6 and Fig. 7-7). It shows a more stable

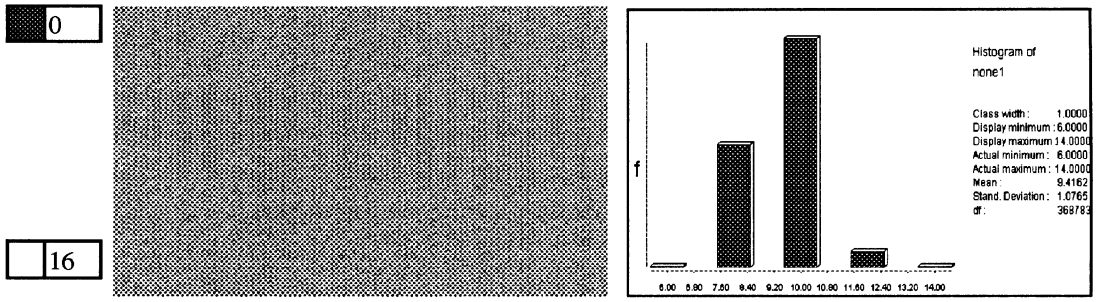
doesn't show the noise pixel. In Fig. 7-7, two groups of black spots on the lower corner of each side are caused by unbalanced illumination, not noise. In this context, the images of the video cameras don't have serious intrinsic noise that should be corrected before target location determination.

On the other hand, in order to compare the resolution power of two cameras, B/W test chart images were captured under the same photographic condition. As shown in Fig. 7-8 and Fig. 7-9, the patterns of the histograms look alike, but the response range of the S-VHS camera(0-235) was wider than that of the VHS camera(0-215).

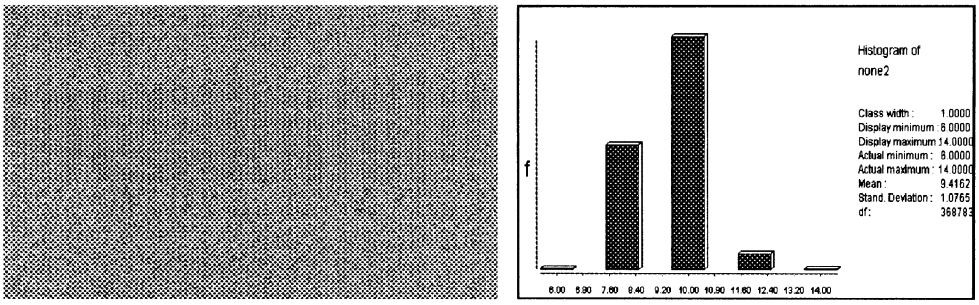
Another comparison was made for the color response characteristics. Three pieces each of red, green, and blue colored sheet were attached to a white background. The size of these colored sheets is same. The results are presented in Fig. 7-10 and Fig. 7-11. White background was viewed as the color of the band itself In all color bands. There are three peaks in the 3 color band histograms. However, there are 4 peaks in the B/W band.

Generally, shutter speed influences the image quality. In order to investigate this influence, a rectangular window consisting of 25*20 pixels was picked out from the black target area, and another same size window was chosen from the white background area . The test images of target and background were captured under indoor florescent light and outdoor sun shine at different shutter speeds. The ranges of gray values for each window were then collected, and plotted in Fig. 7-12. The images captured outdoors show uniform ranges of gray values in spite of varying shutter speed, but the gray values of images captured indoors changed with the shutter speed. The sensitivity of the sensor gradually decreased when the shutter speed increased. Therefore, when photographing indoors with short exposure times(e.g. less than 1/2000s), we should provide enough illumination according to the surrounding conditions.

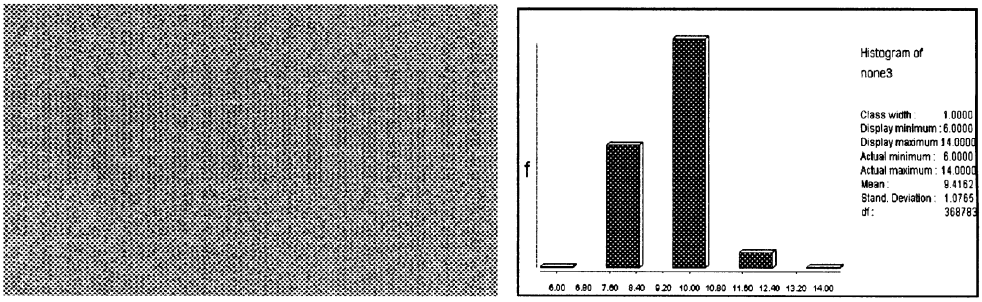
Red Band



Green band



Blue band



B/W Band

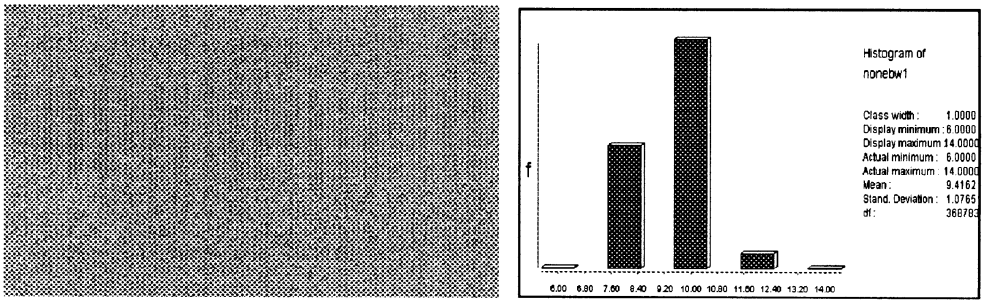
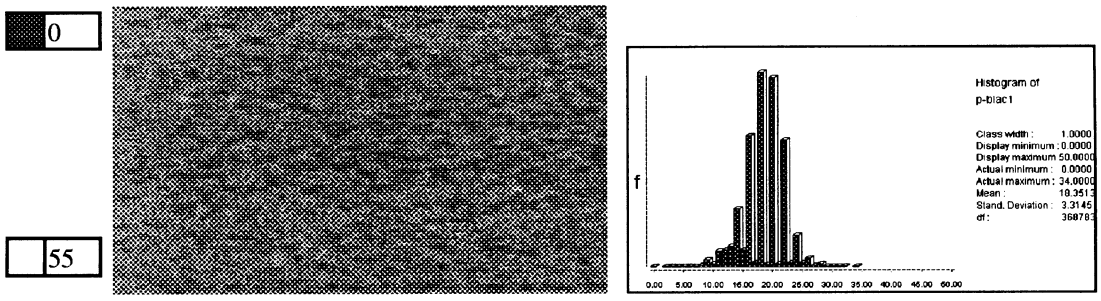
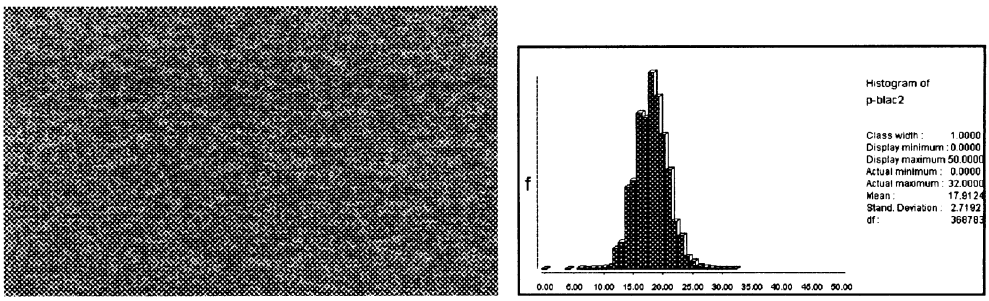


Fig. 7-1. Intrinsic Image of the Frame Grabber.

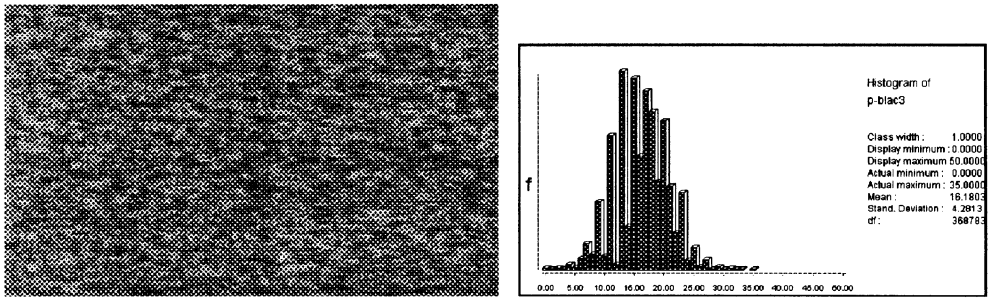
Red Band



Green Band



Blue Band



B/W Band

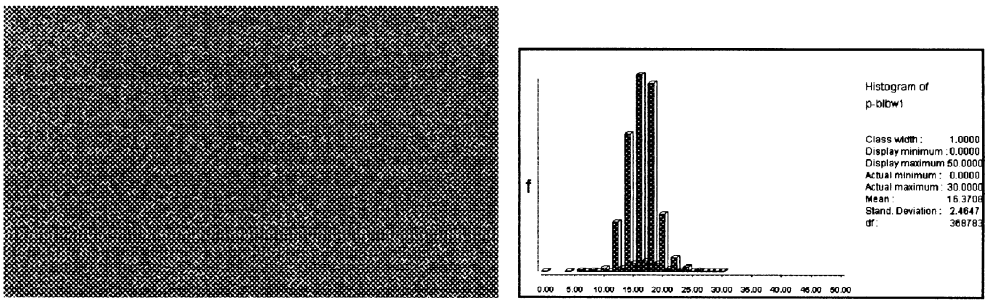
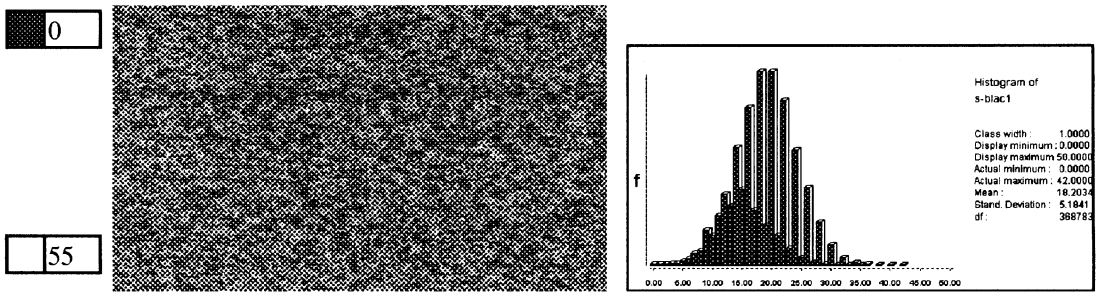
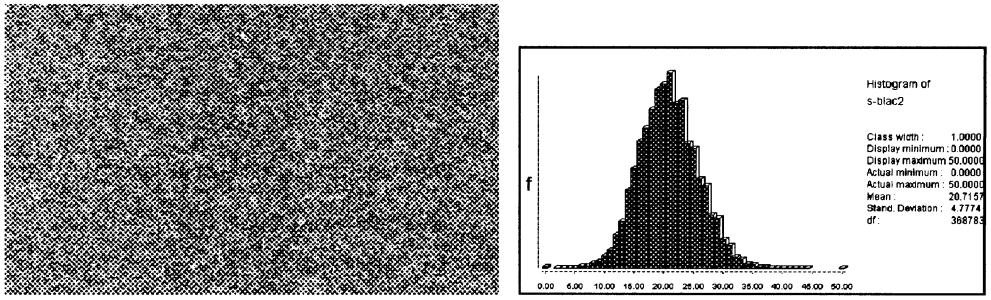


Fig. 7-2. Dark Image (S-VHS Camera and S-VHS VCR)

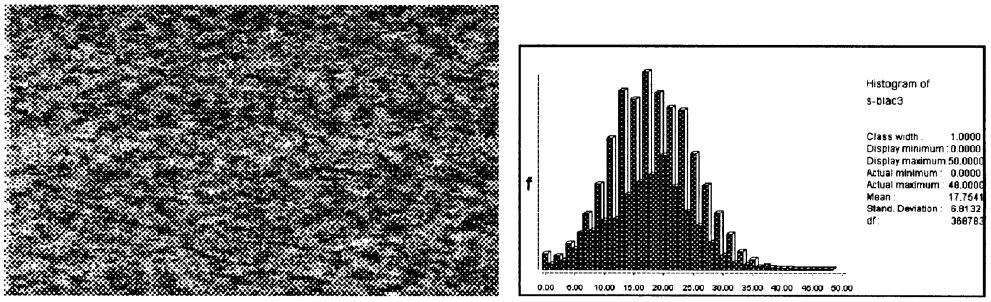
Red Band



Green Band



Blue Band



B/W Band

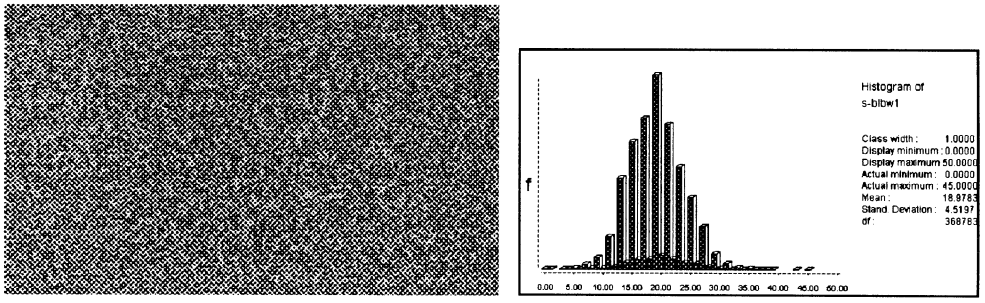


Fig. 7-3. Dark Image (VHS Camera and S-VHS VCR).

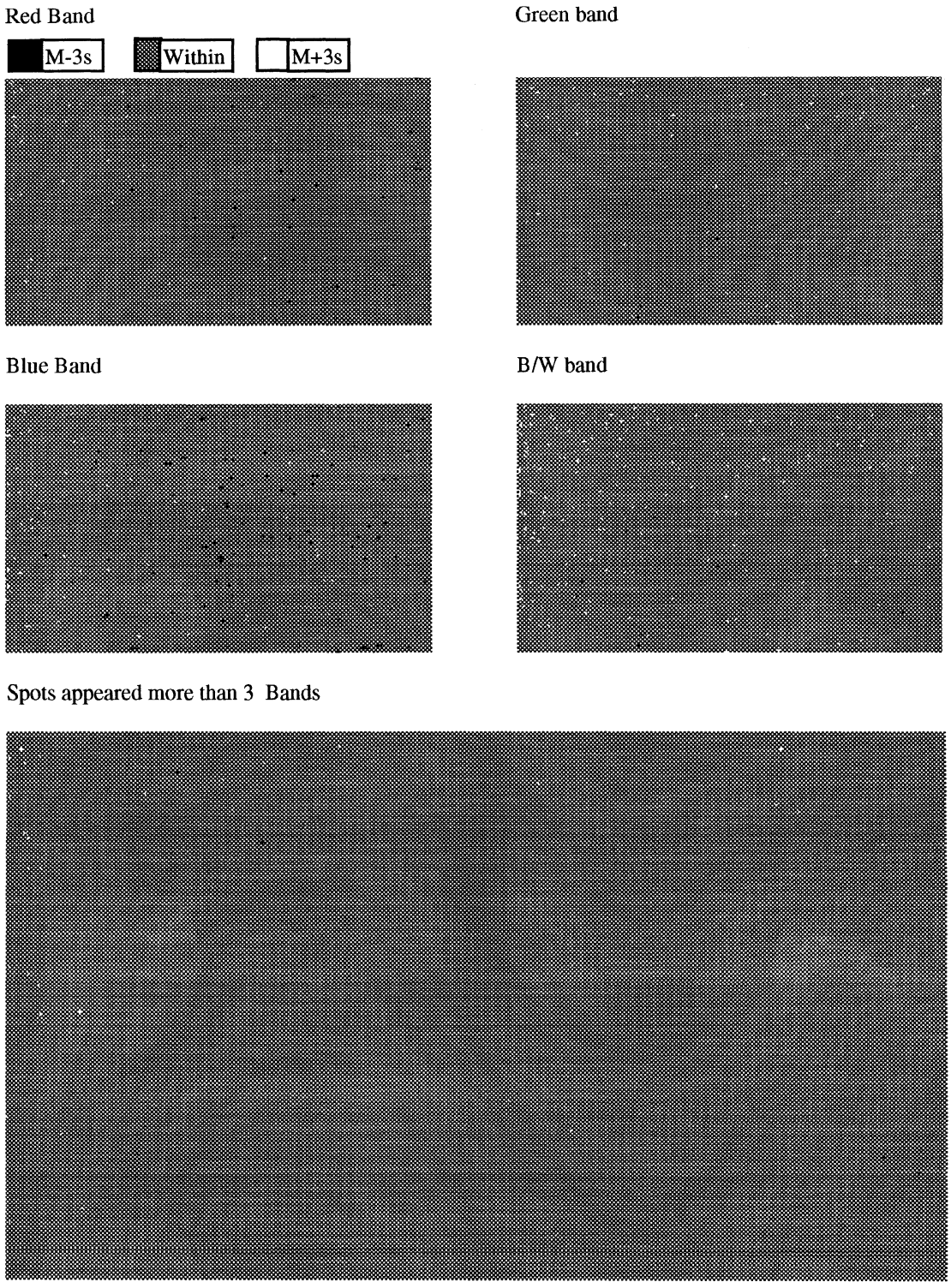


Fig. 7-4. Gross Error Spots of Dark Image (S-VHS Camera and S-VHS VCR)

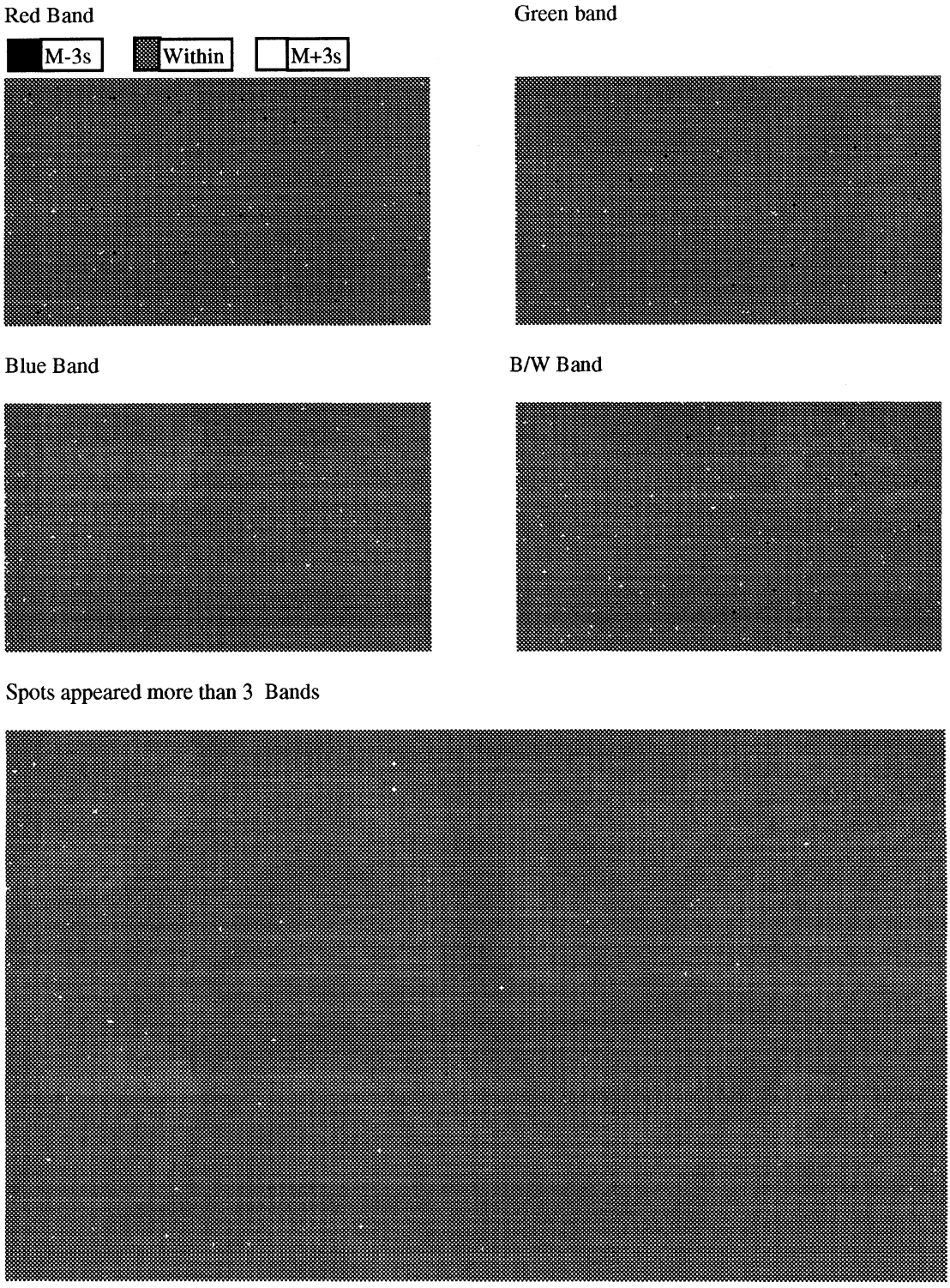
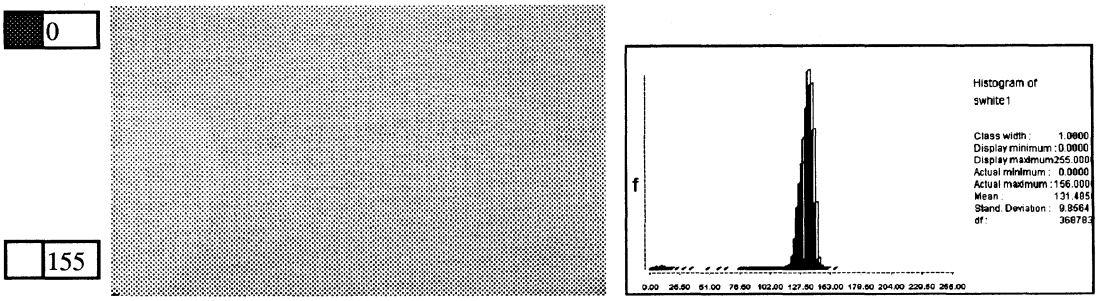
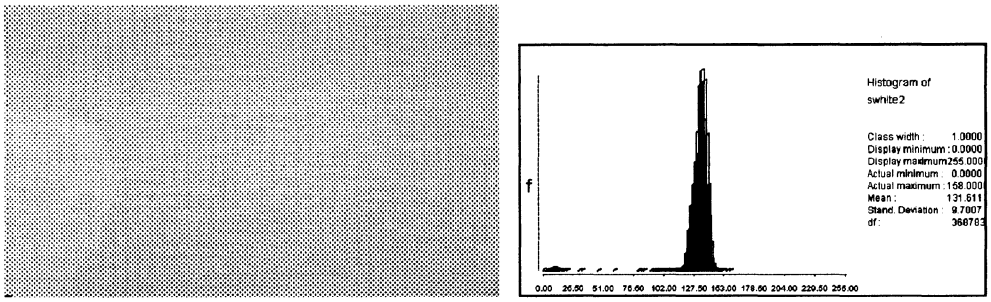


Fig. 7-5. Gross Error Spots of Dark Image(VHS Camera and S-VHS VCR)

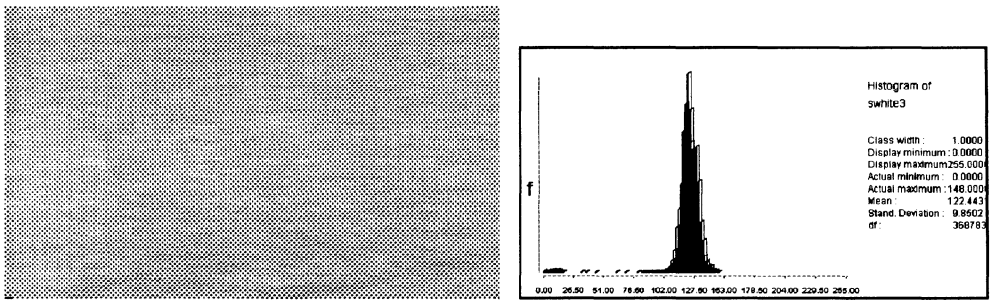
Red Band



Green Band



Blue Band



B/W Band

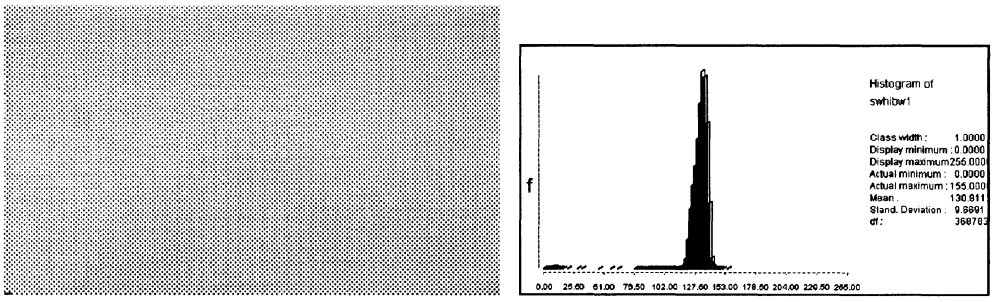


Fig. 7-6. White Image (VHS Camera)

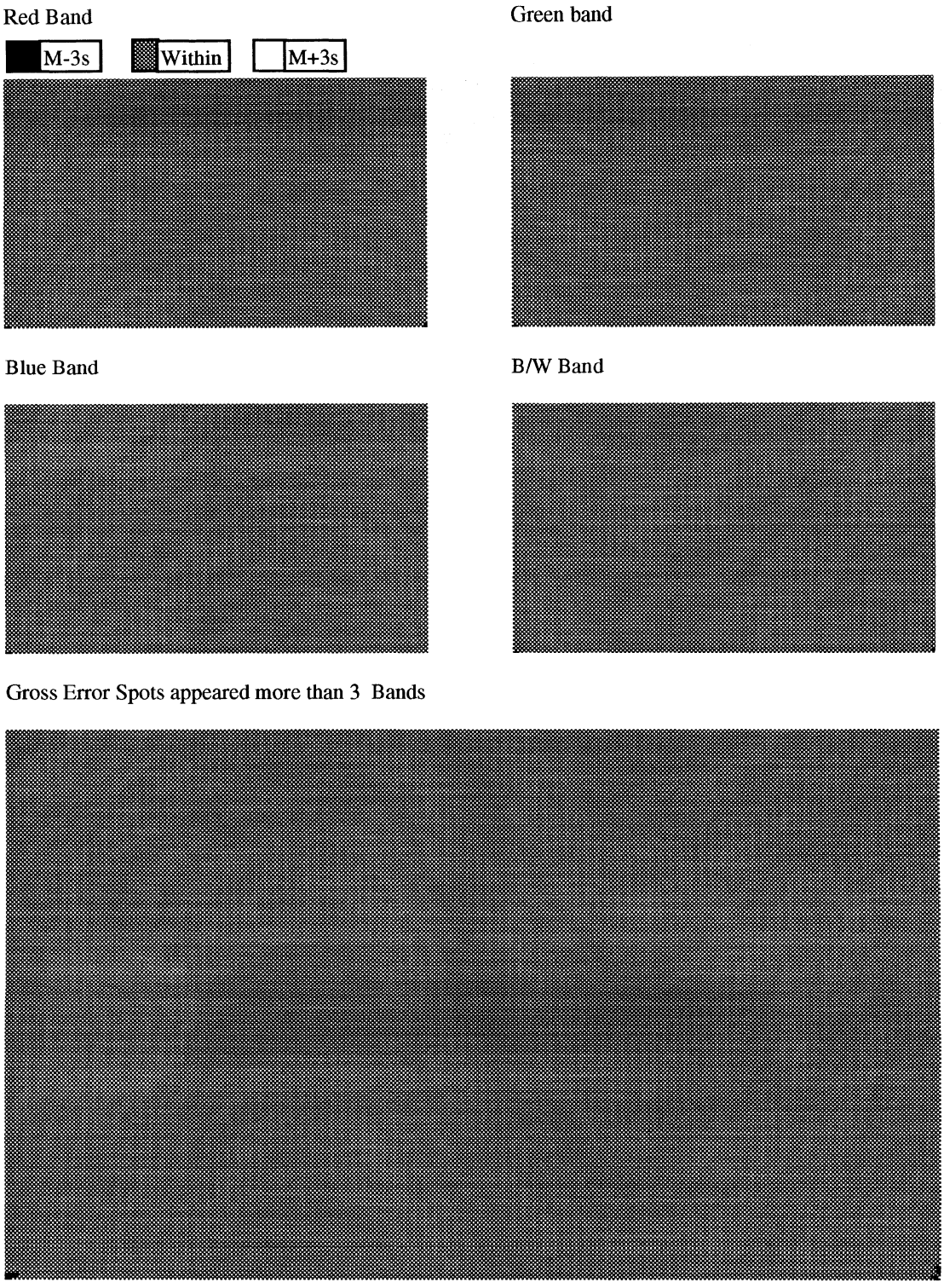


Fig. 7-7. Gross Error Spots of White Image (VHS Camera)

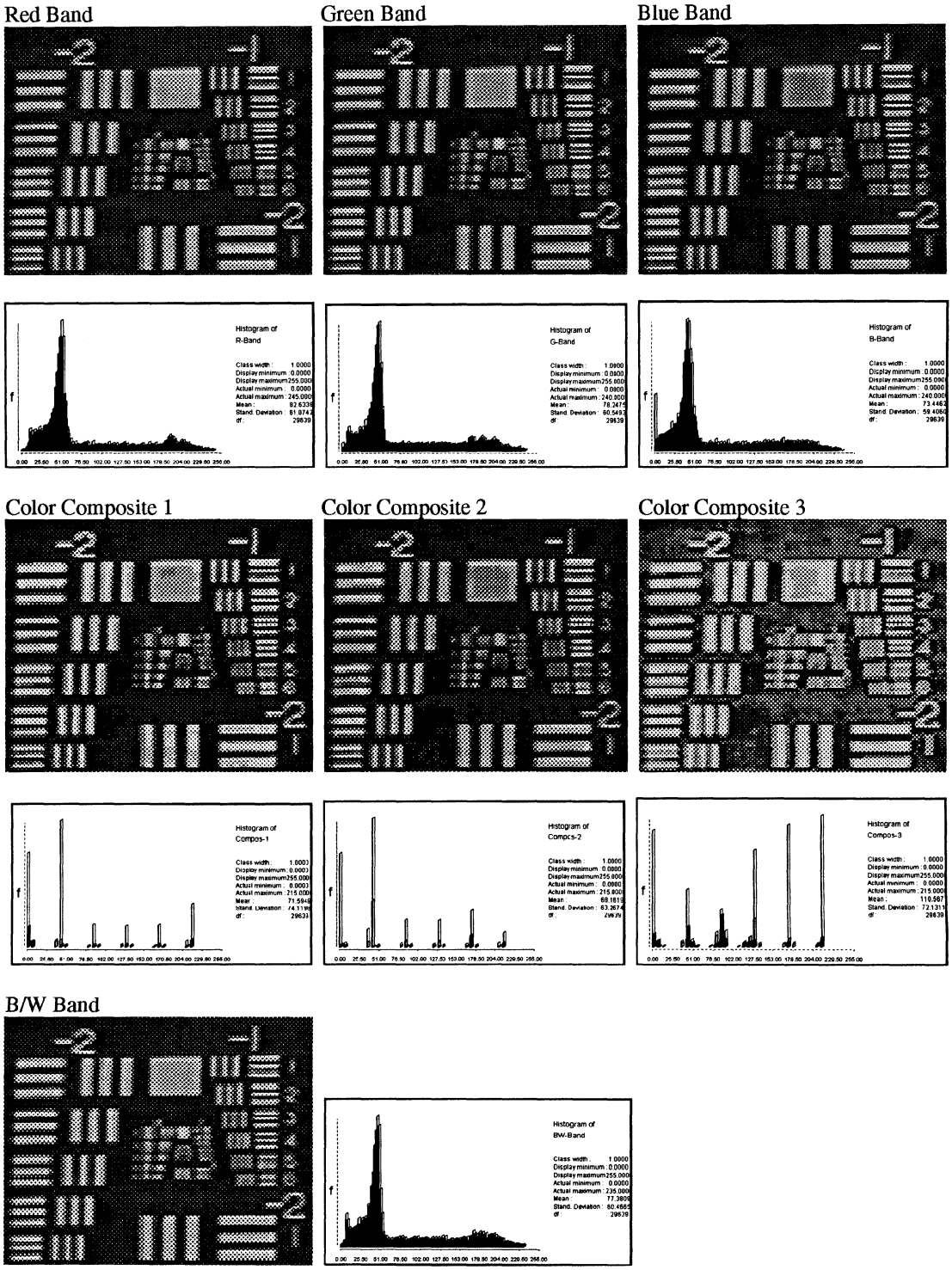


Fig. 7-8. Image of Resolution Power Test Chart (S-VHS Camera)

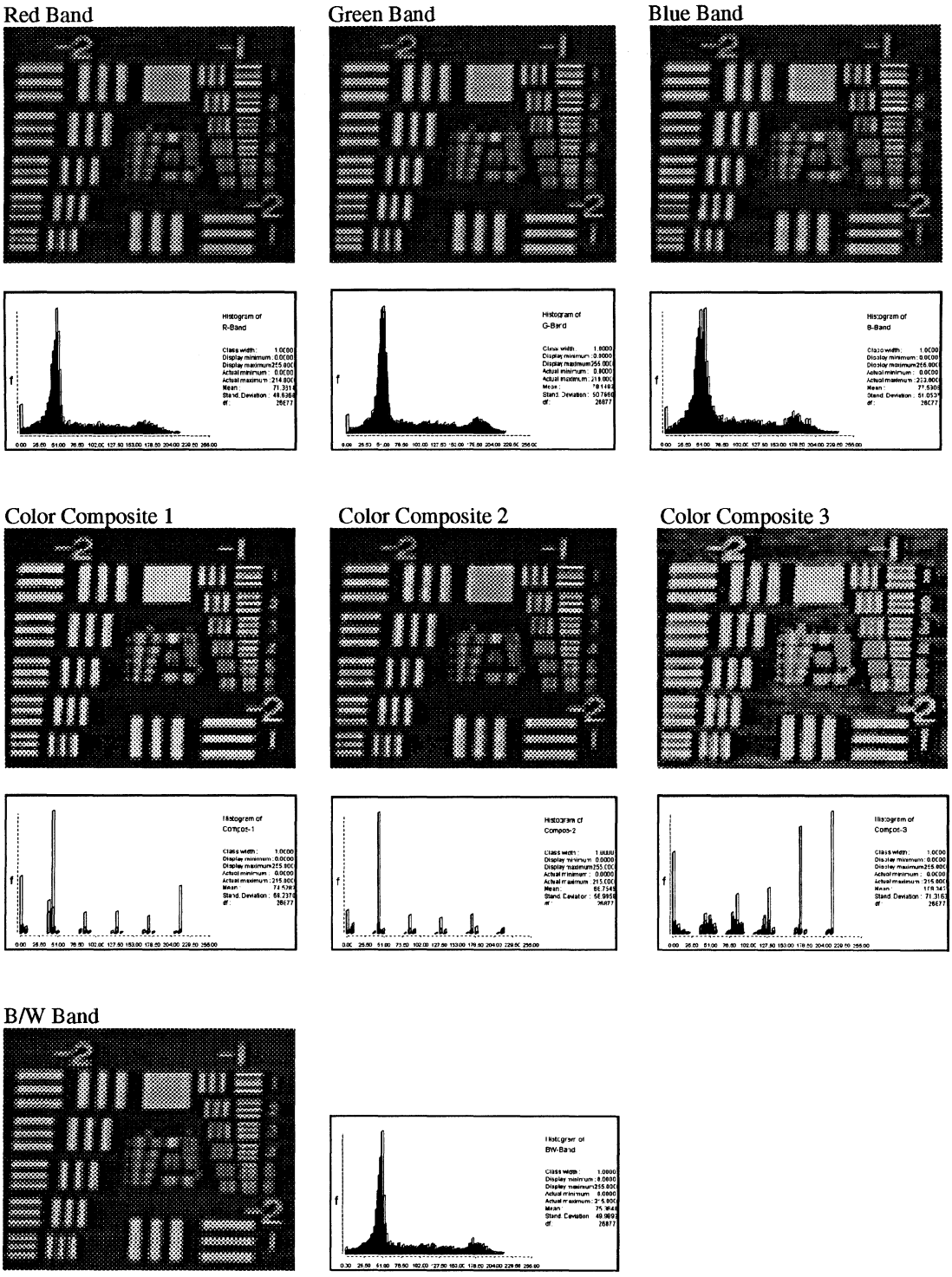


Fig. 7-9. Image of Resolution Power Test Chart (VHS Camera)

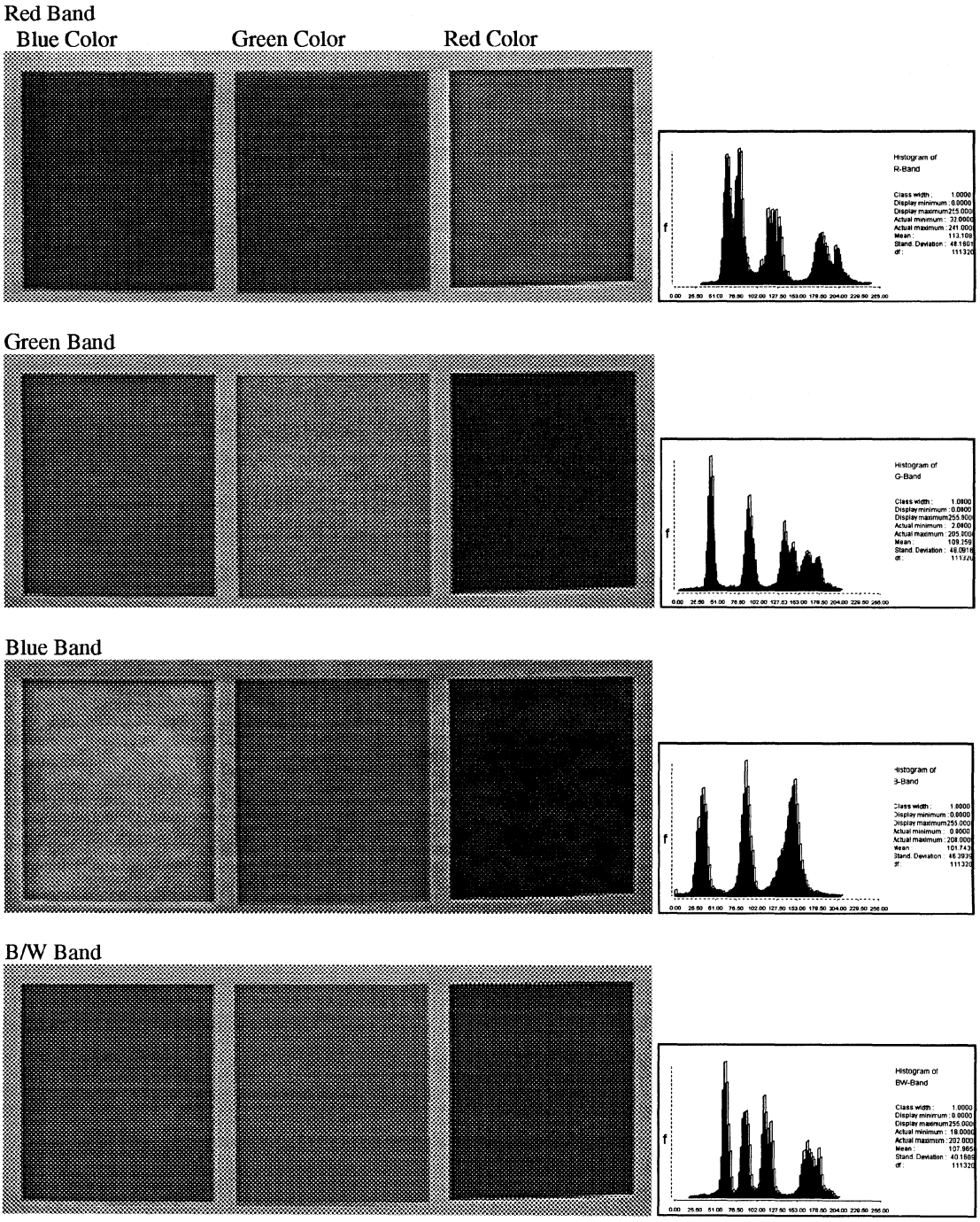


Fig. 7-10. Image of RGB Color Paper (S-VHS Camera and S-VHS VCR)

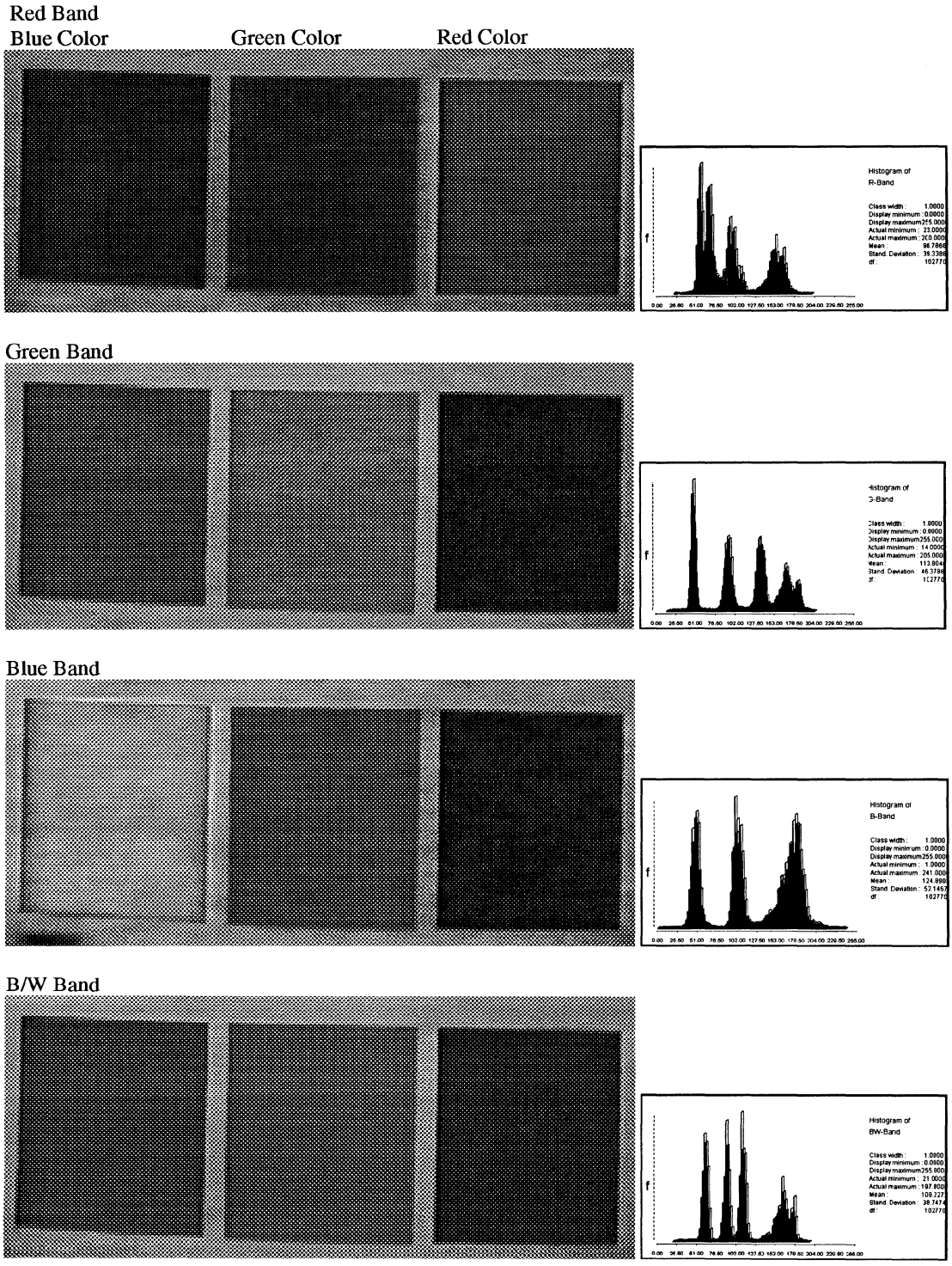


Fig. 7-11. Image of RGB Color Paper (VHS Camera and S-VHS VCR)

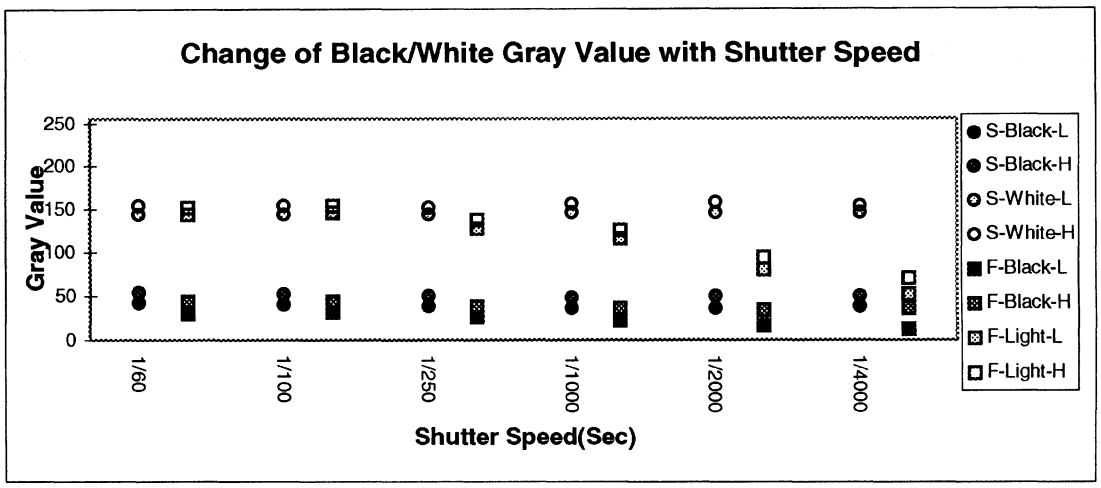


Fig. 7-12. Change of Gray Value with Shutter Speed.

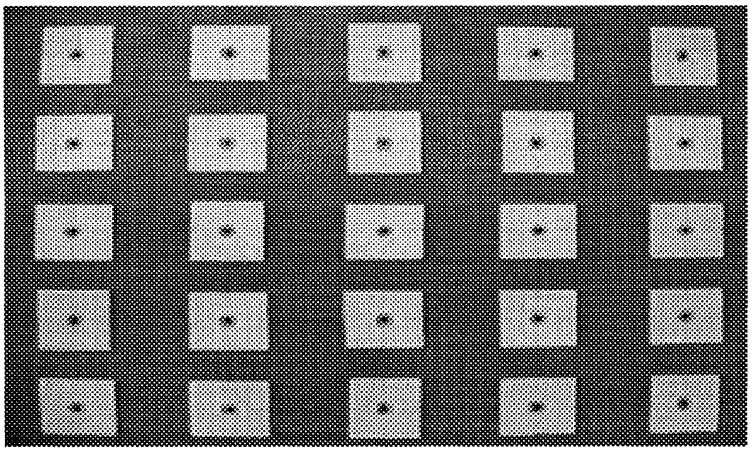
7-1-2. Geometric Characteristics

Compared to metric cameras, non-metric cameras' lenses are designed for high resolution at the expense of geometric quality. Interior orientation parameters lack stability, and there are no fiducial marks. In addition to the above disadvantages, the size of unit pixels of video images changes in the course of digitizing, and the scale changes with the display format. We only have approximate data for the object distance, focal length, and pixel size of the video images. Therefore lens distortions were derived based on some approximations. The horizontal and vertical pixel sizes of the CCD were deduced from dividing the length of each side by the total number of pixels in the corresponding direction.

Fig. 7-13. is an original image of the plane target board. Fig 14 is its processed image. They show typical lens distortion. As seen in Figures 7-15(A) and 7-16(A), the camera has different distortion in vertical, horizontal, and diagonal directions. This means it has radial and decentering distortions. But, it is hard to estimate the amounts of distortion because it was derived, based on

an assumed pixel size. We derived another lens distortion curve showing typical radial distortions by using the calibrated focal length and adjusting the pixel sizes(Figures 7-15(B) and 7-16(B)). The maximum distortion was 6μm for the S-VHS camera and, 3.5μm for the VHS camera which is quite acceptable.

Object Distance : 57cm



Object Distance 67cm

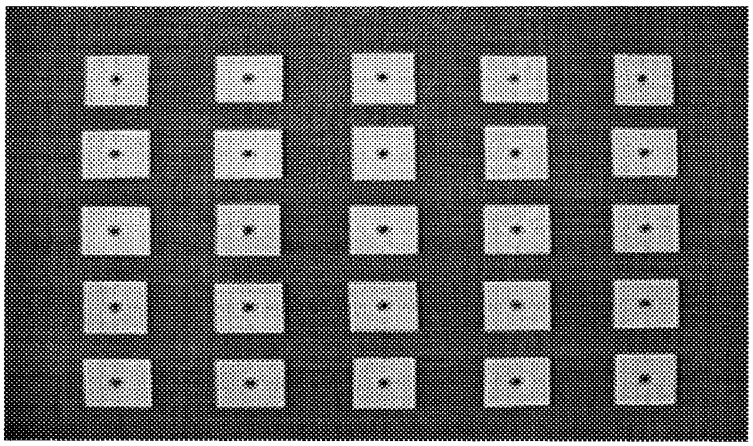


Fig. 7-13. Images of Plane Target Board (Panasonic AG-455P)

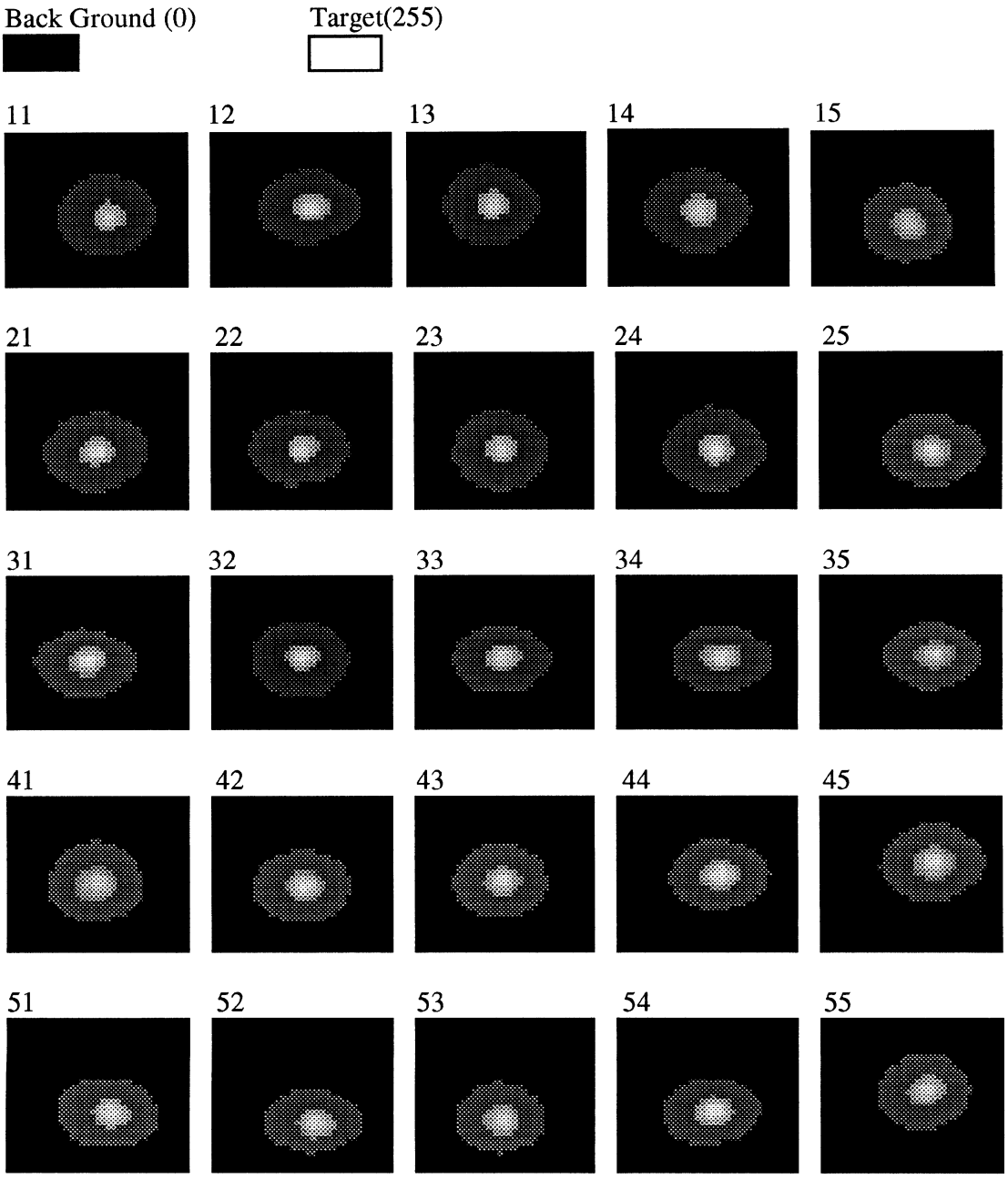
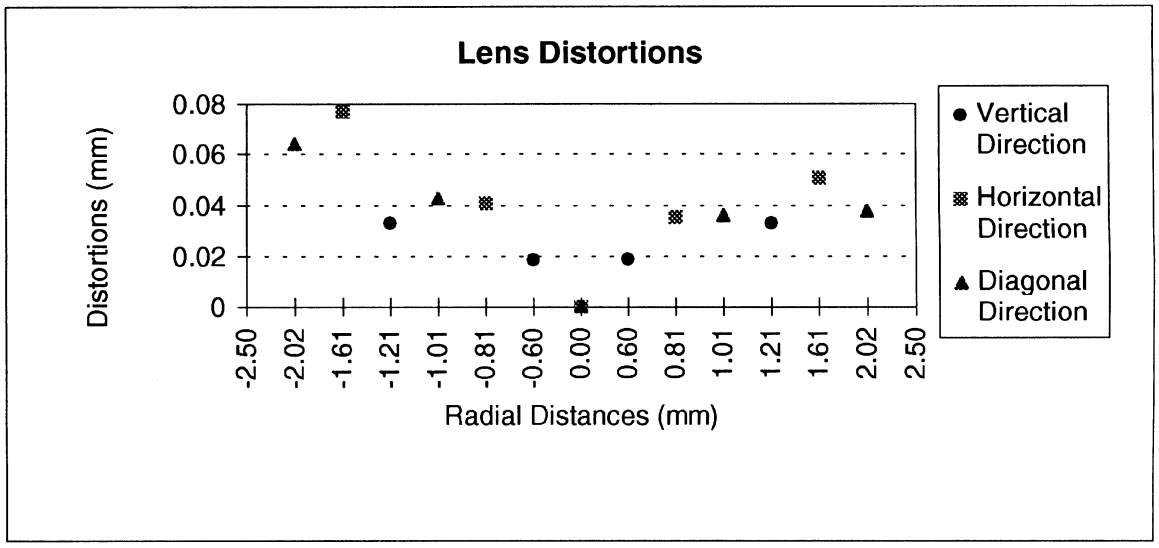


Fig. 7-14. Shapes of Thresholded Targets.

(A) Measured Focal Length



(B) Calibrated Focal Length

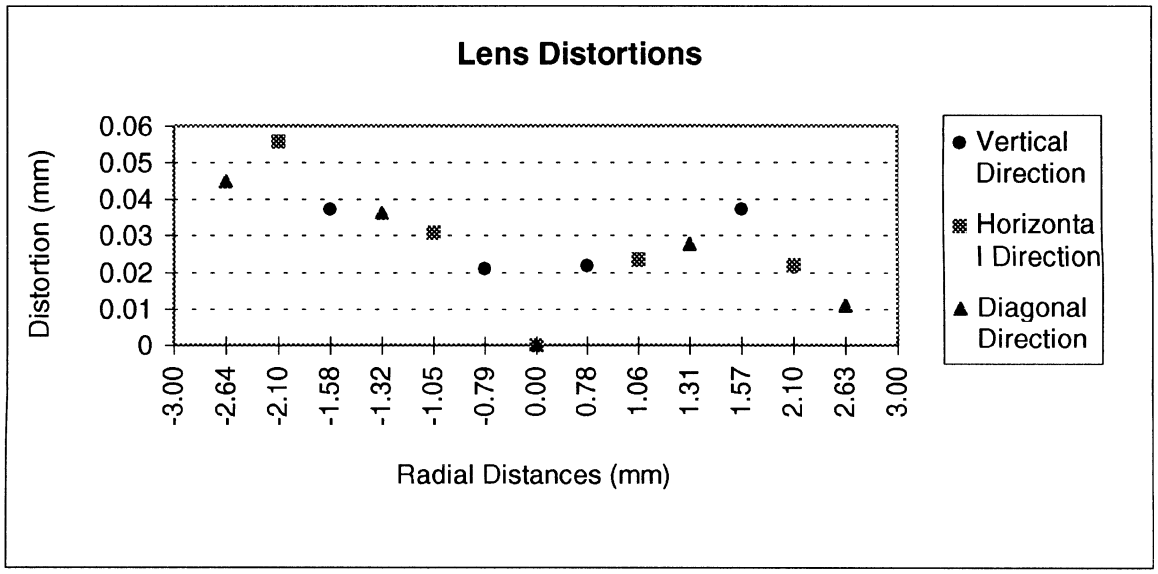
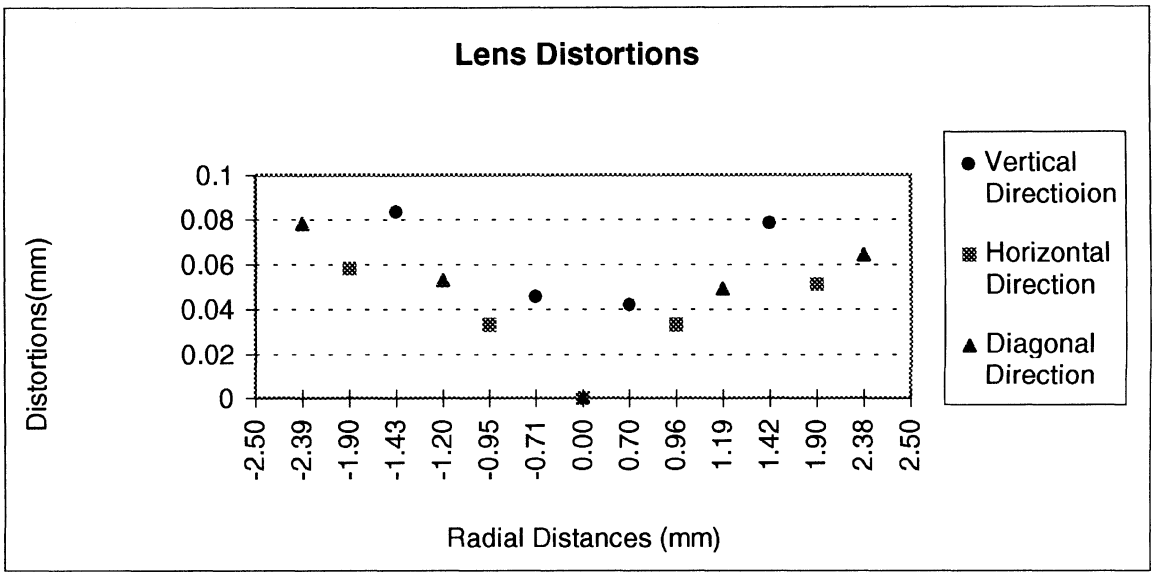


Fig. 7-15. Lens Distortion(Panasonic, AG-455P).

(A) Measured Focal Length



(B) Calibrated Focal Length

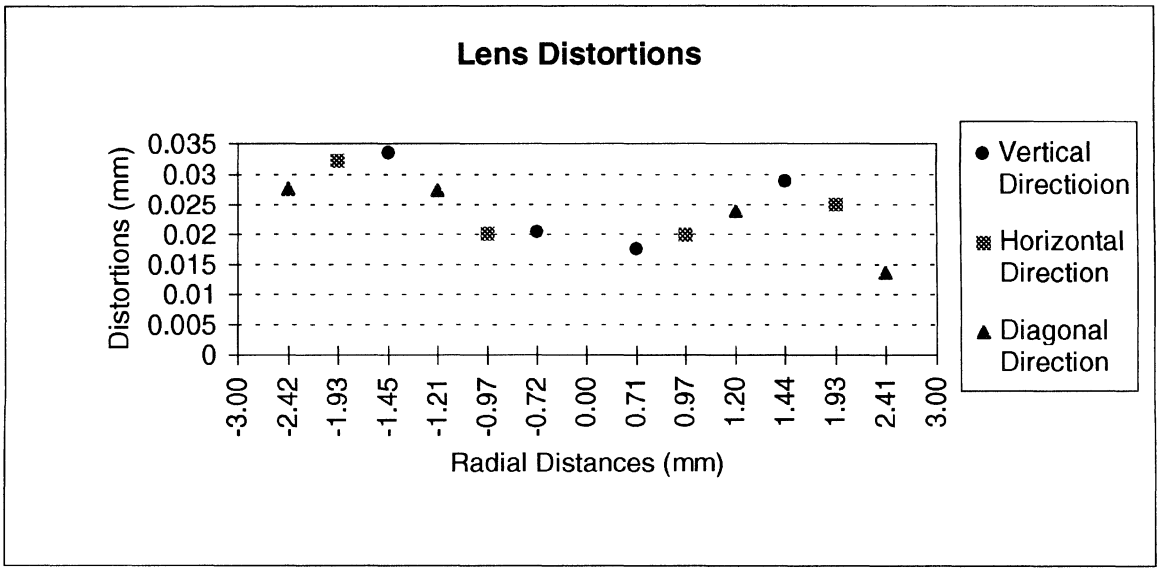


Fig. 7-16. Lens Distortion(Samsung, H-33).

7-1-3. 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. 17(A) fell from 40cm height above the optical axis. Theoretically the speed should be 2.8m/s, but its actual speed was about 1.7m/s. The image blur was improved with the shutter speed. The relative motion during exposure was 30 rows In Fig. 17(A-1), 12 in (A-2), and 3 in (A-3). The target plate shown in the right column in Fig. 17. fell from 60cm height above the optical axis. In this 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 speed.

Vlcek(1988) suggested that the relative motion during exposure should not exceed 0.5 pixels for a clear image for image processing. Although we have not reached 0.5 pixels, we expect that relative motion lower than 0.5 pixels will provide cleaner images. When it comes to dynamic videometry, the lower the relative motion becomes, the higher the measuring accuracy will be. In static videometry, a large image scale usually provides more accurate measurements. However, in images of moving objects, 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.

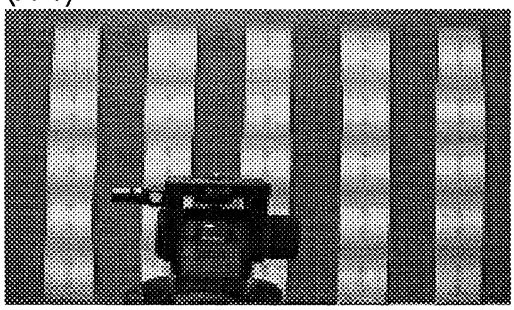
Usually video cameras provide a wide range of shutter speeds from 1/60 s to 1/4000s. However, it should be kept in mind that although the camera is taking a picture within a time interval shorter than 1/60s, it produces just 60 fields in a second(NTSC standard).

Falling Height : 40cm

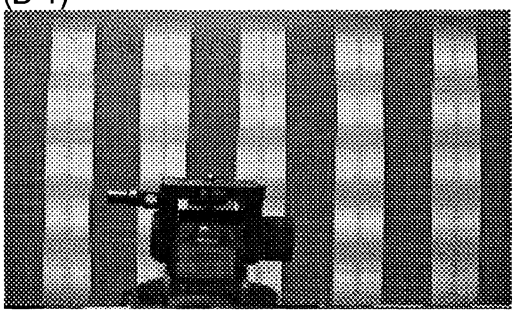
Falling Height : 60cm

Shutter 1/100s

(A-1)

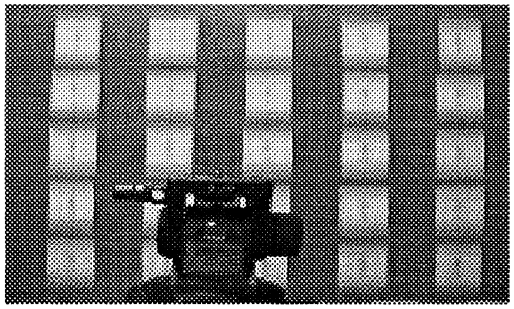


(B-1)

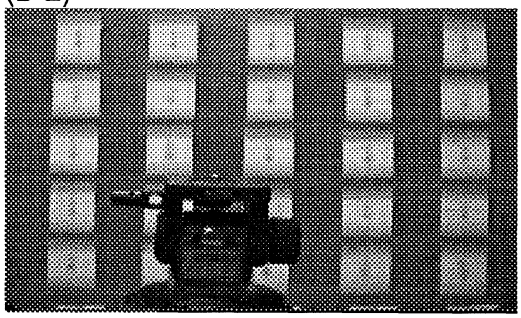


Shutter 1/250s

(A-2)

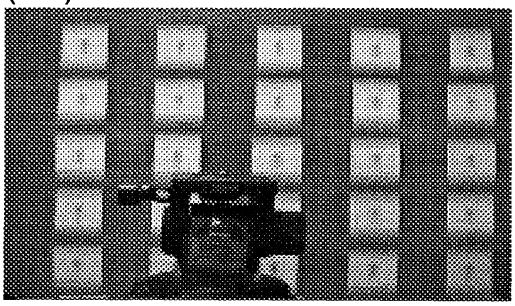


(B-2)



Shutter 1/1000s

(A-3)



(B-3)

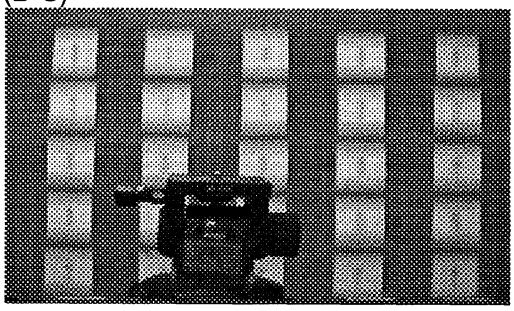


Fig. 7-17. Falling Target Images at Various Shutter Speeds.

7-2. Dynamic Monitoring

7-2-1. 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 using the frame advancing function(1/60s) of the VCR. Before capturing the real sequential image, significant scenes(e.g. driver's motion, lights on and off, door open and close) were designated as milestones. We counted the number of fields between each milestone several times. It does not exceed 2 field's(1/30s) difference during 10s. In this context, we captured every 4^{th} field(1/15s), comprising synchronized sequential images for monitoring the movement of the car. The PC works on an 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, with a file size of 1.1Mb. This image is a 24-bit Band-Interleaved-Pixel(BIP) format, and was converted into Band Sequential(BSQ) images in the "IDRISI for Windows" format. There were no significant image quality differences between the S-VHS and VHS video cameras when the images were recorded by the S-VHS VCRs using the S-Video format. 77 pairs of sequential images were digitized from S-VHS VCR. Fig 7-18. shows a pair of left and right images. Image coordinates for three points in each image were processed to determine their object coordinates.

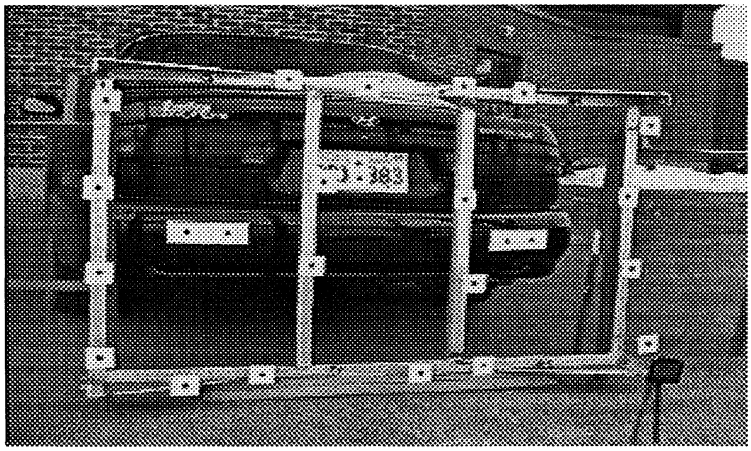
7-2-2. Sub-Pixel Target Coordinates

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

- Noise removal using median filter.
- Searching window centered on target.
- Reclass based on a threshold.
- Determination of the target center coordinates using equation(3-4).

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

(A) Left Image



(B) Right Image

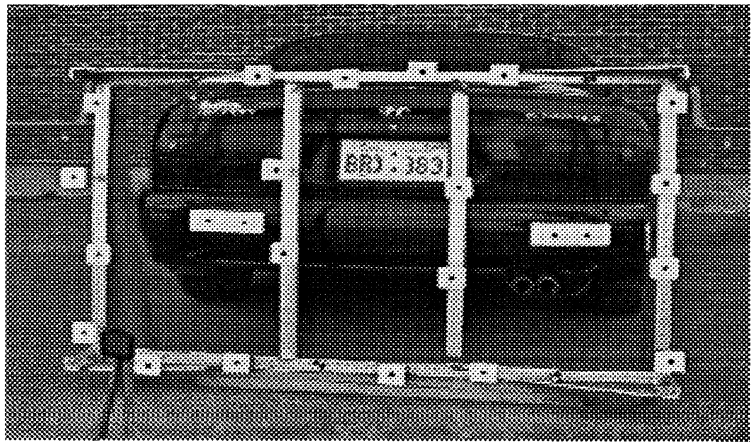


Fig. 7-18. Left and Right Images of the Moving Car(B/D=0.46).

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. Fig. 19(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 was nearly zero, and the maximum difference was 0.02 pixel in 20 targets.

7-2-3. 3-D Object Coordinates

For 3-dimensional tracking of the targets in each image sequence, we prepared image coordinates of 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

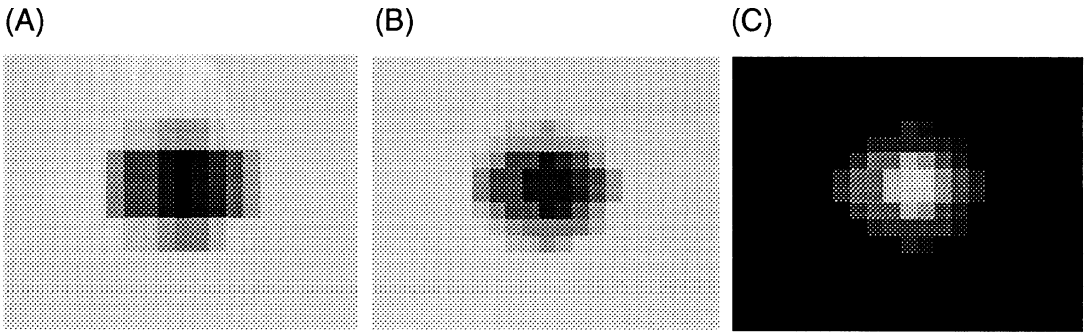


Fig. 7-19. Images of Circular Target in Various Processing Stages.

work from obtaining image coordinates to running the program. As described in Chap. 4, DLT and UNBASC2 could be used to derive 3-dimensional object coordinates using non-metric image coordinates.

While DLT affords pixel units as a 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, and image coordinates were computed using equation(3-1). 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. Fig. 20. is a flow chart explaining the overall procedure to obtain 3-Dimensional coordinates of targets.

It is well known that the imaging geometry plays a great role in determining the accuracy. When the imaging geometry is ill conditioned, the accuracy of DLT becomes worse when especially increasing the number of parameters.

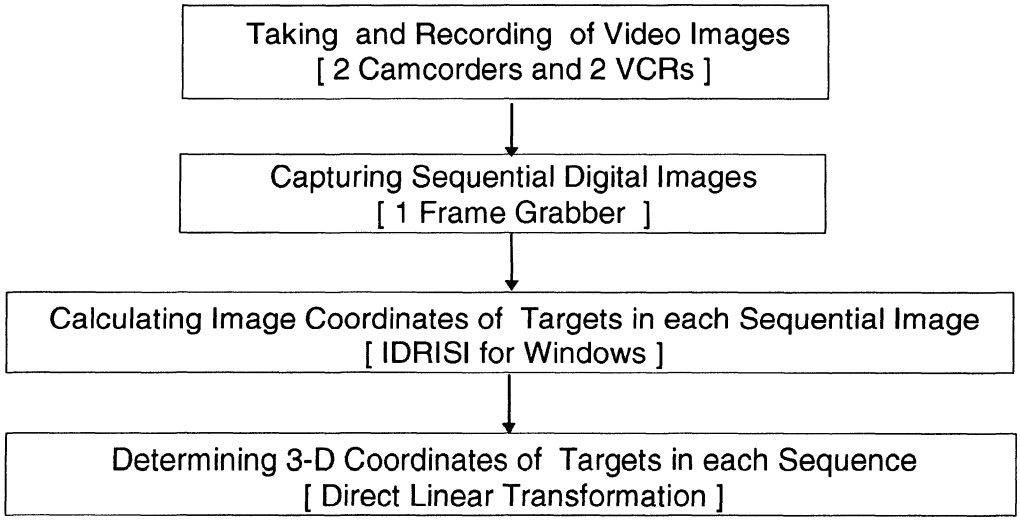


Fig. 7-20. Procedure to Obtain 3-Dimensional Object Coordinates.

Fig. 7-21 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, who showed the largest RMS error was sequentially excluded from the control points. At last, optimum control values were obtained except for 2 points.

This optimum control point set provides a little smaller RMS error when the unknown parameters of DLT were increased. However, the RMS errors of the unknown points were minimum when the unknown parameters were 12. Therefore, 12 parameters were adopted to determine the tracking of targets attached to the moving car. As shown in Table 7-1 , 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 1mm. Generally, the accuracy between rows(y axis)

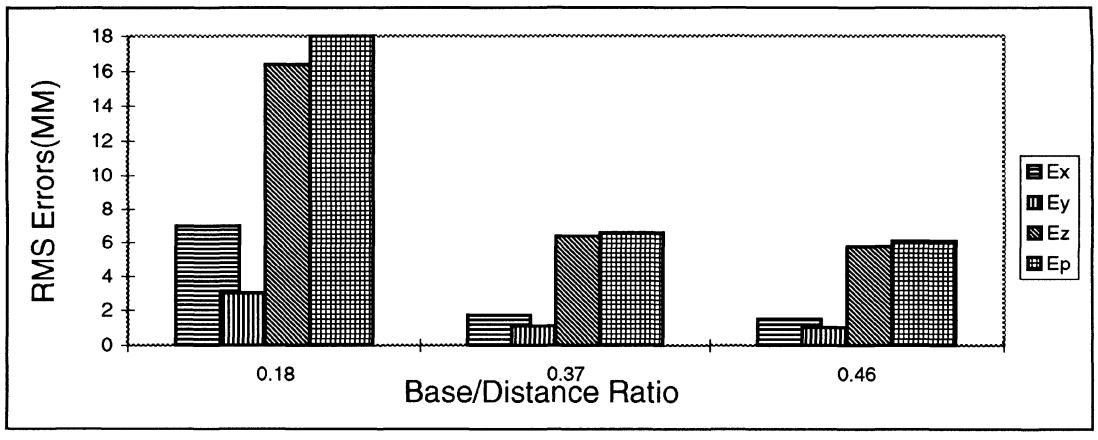


Fig. 7-21. RMS Errors of DLT for Different B/D(Unknown Parameters: 12)

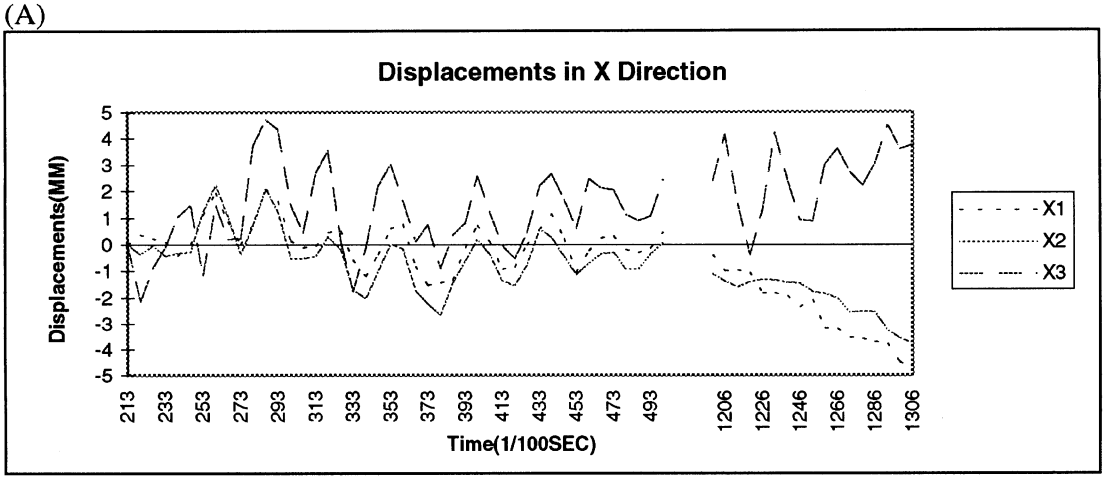
Table 7-1. RMS Errors of Object Coordinates(Unknown Parameters of DLT:12)

	Ex(mm)	Ey(mm)	Ez(mm)	Ep(mm)
Control Points(18)	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

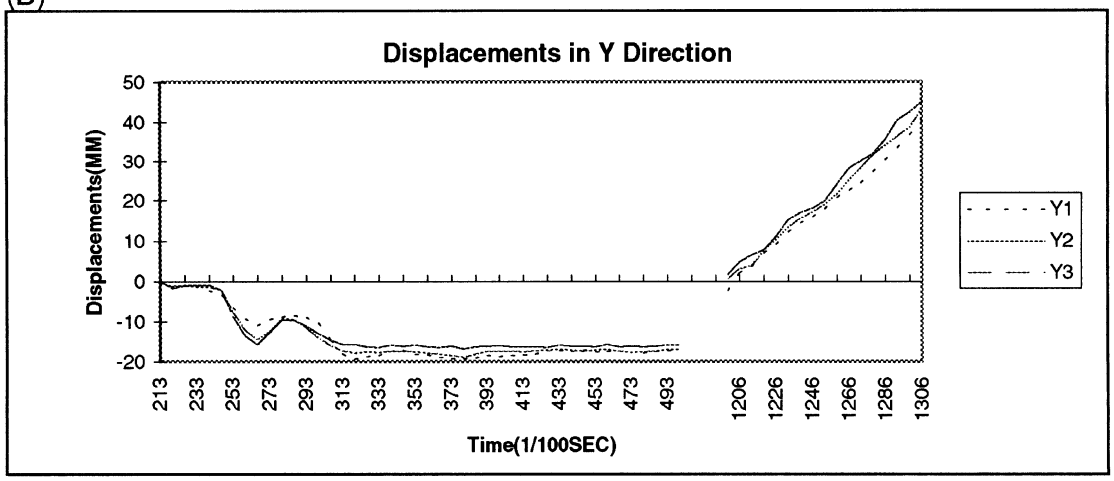
is much better than that of columns in the video images. RMS errors in x direction of targets 1 and 2 are about 1.5 - 3 times those in y direction. But target 2 shows a reverse phenomenon.

At last, we got the trajectory of targets attached to the car. In Fig. 7-22(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, they gradually become weakened. The second half of diagram describes the trajectory when the car starts to drive ahead. Target 3 was attached to a rounded surface, and its narrow intersection angle between two cameras might cause some large deviation of the trajectory 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 clearly. In z direction, it remains still while the driver gets into the car and starts the ignition. When the car starts to drive ahead, the targets' trajectories grow farther.

(A)



(B)



(C)

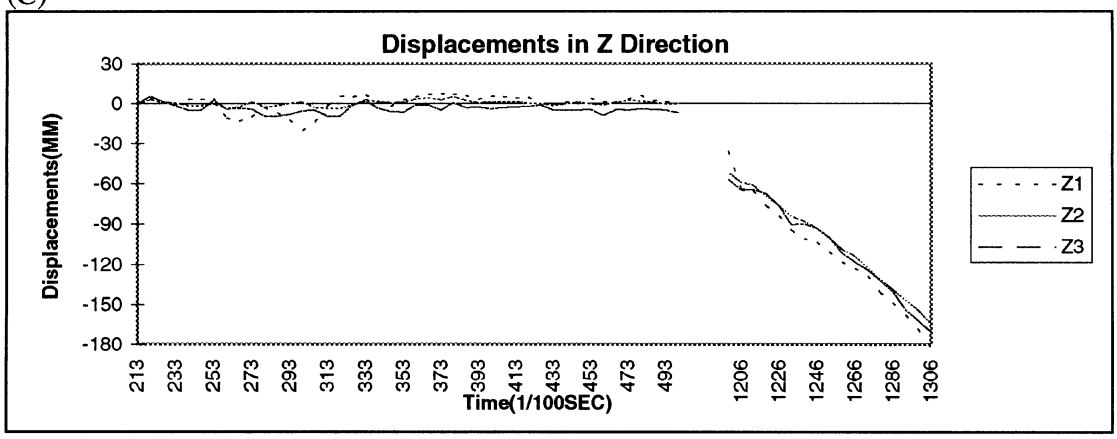


Fig. 7-22. Trajectories of Targets Attached to a Moving Car

8. CONCULSIONS

Two video cameras and VCR combined with frame grabber were investigated and tested for vibration and movement monitoring.

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. However, the TV broadcasting standards prevent a better high resolution video image. It will be improved with the advent of HD-TV.

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 synchronization in this experiment, more elaborate electronic devices are indispensable to monitor high frequency vibrations. Especially, a single camera system(e.g. mirror photogrammetry and beam splitter photogrammetry) which simultaneously records right and left images in a single image is highly recommend for this purpose.

Image processing software for the PC(e.g. IDRISI) was a usable tool for image preprocessing and target centering. For the real time photogrammetric 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. For a real time video system, it must be improved to get the object coordinates within 1/60s.

In conclusion, a video system and appropriate photogrammetric principles provided the expected results for monitoring of a moving car.

REFERENCES

Abdel-Aziz, Y.I. and Karara, H.M. (1971). "Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close Range Photogrammetry." Proceedings of the AUI/UI Symposium on Close-Range Photogrammetry, Urbana, 1-18.

Abdel-Aziz, Y.I. and Karara, H.M. (1974). "Photogrammetric Potentials of Non-metric Cameras." Civil Engineering Studies, Photogrammetry Series No. 36, University of Illinois, Urbana.

Ackermann, F. (1984), "Digital Image Correlation: Performance and Potential Application in Photogrammetry." Photogrammetric Record. 11(64): 429-439.

Baarda, W. (1967), "A Testing Procedure for Use in Geodetic Networks." Netherlands Geodetic Commission. Publication on Geodesy, Vol. 2, No. 4, Delft.

Baltsavias, E.P. & Stallmann, D. (1991), "Trinocular Vision System for Automatic and Robust Three-Dimensional Detection of the Trajectories of Moving Objects. " Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 8, 1079-1086.

Beyer, H. A. (1992), "Advances in Characterization and Calibration of Digital Imaging Systems." International Archives of Photogrammetry and Remote Sensing, Zurich, Vol. 29, Comm. 5, 545-555.

Beyer, H. A. (1995), "Automated Dimensional Inspection with Real-Time Photogrammetry." ISPRS Journal of Photogrammetry and Remote Sensing, 50(3):20-26.

Bopp, H. & Krauss, H. (1977), "A Simple and Rapidly Converging Orientation and Calibration Method for Non-Topographic Applications." Proceedings of the American Society of Photogrammetry, Fall Technical Meeting, 425-432.

Curry , S., Baumrind, S. & Anderson, J. M. (1986), "Calibration of an Array Camera." Photogrammetric Engineering & Remote Sensing, Vol. 52, No 5, 627-636.

Dermanis, A. (1994), "Free Network Solutions with the DLT Method." ISPRS Journal of Photogrammetry and Remote Sensing, 49(2):2-12.

Dunbar, P. (1986), "Machine Vision." Byte, January 1986, 161-173.

EL-Hakim, S. F. (1986), "Real-Time Image Metrology with CCD Cameras," Photogrammetric Engineering & Remote Sensing, Vol. 52, No. 11, 1757-1766.

Faig, W. (1972), "Single Camera Approach in Close Range Photogrammetry." Proceeding of the 38th Annual Meeting, American Society of Photogrammetry, Washington, D. C., 1-8.

Faig, W. (1975a), "Photogrammetric Equipment Systems with Non-Metric Cameras." Close-Range Photogrammetric Systems, American Society of Photogrammetry, Falls Church, Virginia, 648-657.

Faig, W. (1975b), "Calibration of Close-Range Photogrammetric Systems: Mathematical Formulation." Photogrammetric Engineering & Remote Sensing, Vol. 41, No. 12, 1479-1486.

Faig, W. & Shih, T. Y. (1986), "Critical Configuration of Object Space Control Points for Direct Linear Transformations." International Archives of Photogrammetry and Remote Sensing, Vol. 26, Comm. 5, 23-29.

Faig, W., El-Habrouk, H., Li, X.P., and Hosny, M. (1996), "A Comparison of the Performance of Digital and Conventional Non-metric Cameras for Engineering Applications." International Archives of Photogrammetry and Remote Sensing, Vol. 31, B5, 147-151.

Fraser, C. S. (1982), "On the Use of Nonmetric Cameras in Analytical Close-Range Photogrammetry." The Canadian Surveyor, Vol. 36, No. 3, 259-279.

Granshaw, S.I. (1980), "Bundle Adjustment Methods in Engineering Photogrammetry." Photogrammetric Record, 10(56): 181-207.

Grün, A. (1978), "Accuracy, Reliability and Statistics in Close-Range Photogrammetry." International Archives of Photogrammetry, Stockholm, Vol. 22, Comm. 5, 24 pages.

Gruen, A. (1988), "Towards Real-Time Photogrammetry", Photogrammetria, Vol. 42, 209-244.

Gruner, H. (1955), "New Aspects of Mono-Photogrammetry." Photogrammetric Engineering, Vol. 21, No. 1, 38-49.

Heipke, C. Stephani, M., Strunz, G. & Lenz, R. (1992), "Photogrammetric Calibration and Point Determination Using a Digital CCD Camera." International Archives of Photogrammetry and Remote Sensing, Washington D. C., Vol. 29, Comm. 5, 556-560.

Huang, T. S. & Tsai, R. Y. (1981), "Image Sequence Analysis: Motion Estimation." Image Sequence Analysis(Editor: T.S. Huang), Springer-Verlag, 1-18.

Karara, H. M.(1980), "Non-Metric Cameras." Development in Close Range Photogrammetry-1(Edited by Atkinson), Applied Science Publishers, 63-80.

Laurin, D. G. (1993), "A Videometric System for a Flexible Space Structure Emulator." ISPRS Journal of Photogrammetry and Remote Sensing, 48(4):2-11.

Lee, C. K. & Faig, W. (1996), "Vibration Monitoring with Video Cameras." International Archives of Photogrammetry and Remote Sensing, Vienna, Vol. 31, B5, 152-159.

Lenz, R. (1989), "Image Data Acquisition with CCD Camera." Conference on Optical 3-D Measurement Techniques, Vienna, 22-34.

Lenz, R. & Fritsch, D. (1990), "Accuracy of Videometry with CCD Sensors." ISPRS Journal of Photogrammetry and Remote Sensing, 45(2): 90-110.

Lusch, David P. (1988), "Super VHS for Improved Airborne CIR Videography." First Workshop on Videography, ASPRS, Terre Haute, Indiana, 23-29.

Lu, J. , Lin, Z. & Pan, D. (1992), "Single CCD Camera based Three Dimensional Measurement System for Moving Object." International Archives of Photogrammetry and Remote Sensing, Washington, D. C., Vol. 29, Comm. 5, 469-474.

Marzan, C. T. & Karara, H. M. (1975), "A Computer Program for Direct Linear Transformation Solution of the Colinearity Condition, and Some Applications of it." Symposium on Close-Range Photogrammetric Systems, Champaign, Illnois, American Society of Photogrammetry, 420-471.

McGlone, J. C. (1989), "Analytic Data-Reduction Schemes in Non-Topographic Photogrammetry." Non-Topographic Photogrammetry (2nd ed.), American Society for Photogrammetry and Remote Sensing, 37-57.

Meisner, Douglas E. (1986), "Fundamentals of Air Borne Video Remote Sensing." Remote Sensing of the Environment, Vol. 19, 63-79.

Mikhail, E. M., Akey, M. L. & Mitchel, O. R. (1984), "Detection and Sub-pixel Location of Photogrammetric Targets in Digital Images." Photogrammetria, Vol.39, 63-83.

Moniwa, H (1977), "Analytical Photogrammetric System with Self-Calibration and its Applications." Ph.D. Dissertation, University of New Brunswick, Fredericton, NB, Canada.

Naftel, A.J. & Boot, J.C. (1991), "An Iterative linear Transformation Algorithm for Solution of the Collinearity Equations." Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 7, 913-919.

Ruther, H. & Parkyn, N. (1990), "Near Real Time Photogrammetry on a Personal Computer." Photogrammetric Record, 13(75):415-422.

Schenk, T. & Hoffmann, (1986), "Stereo Matching Using Line Segments of Zero-Crossings." Proceedings of International Society for Photogrammetry and Remote Sensing, Symposium, Comm. III, Vol. 19, 362-368.

Schenk, T., Li, J. and Toth C. (1991), "Towards an Autonomous System for Orienting Digital Stereopairs." Photogrammetric Engineering & Remote Sensing, Vol. 57, No. 8, 1057-1064.

Tabatabai, A.J. and Mitchell, O.R. (1984), "Edge Location to Sub-Pixel Values in Digital Imagery." IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-6, No.2, 188-201.

Trinder, J. C. (1989), Precision of Digital target Location. Photogrammetric Engineering & Remote Sensing, Vol. 55, No. 6, 883-886.

Trinder, J. C., Jansa, J. & Huang, Y. (1995), "An Assessment of the Precision and Accuracy of Methods of Digital Target Location." ISPRS Journal of Photogrammetry and Remote Sensing, 50(2):12-20.

van Wijk, M. C. & Ziemann, H. (1976), "The Use of Non-Metric cameras in Monitoring High Speed Process." Photogrammetric Engineering & Remote Sensing, Vol. 42, No.1, 91-102.

Vlcek, J. (1988), "Nature of Video Images." First workshop on videography, ASPRS, Terre Haute, Indiana, 5-12.

Vosselman, G. & Förstner, W. (1988), "The Precision of a Digital Camera." International Archives of Photogrammetry and Remote Sensing, Vol. 27, Comm. 1-B, 148-157.

Wong, K. W. & Ho, W.-H. (1986), "Close-Range Mapping with a Solid State Camera." Photogrammetric Engineering & Remote Sensing, Vol. 52, No.1, 67-74.