A COMPUTER VISION SYSTEM FOR MAPPING HUMAN BODIES

Dr. Kam W. Wong, Professor of Civil Engineering
Ying Ke, Research Assistant
Dr. Mary Slaughter, Assoc. Professor of Kinesiology
University of Illinois at Urbana-Champaign
U.S.A.
Commission V

Dr. Randall Gretebeck
Johnson Space Center, NASA
U.S.A.
Commission V

ABSTRACT:

A computer vision system was developed to map the surface of an entire human body in a highly automatic manner. The system was intended to be used for measuring a large number of human subjects to provide data for studying growth pattern, as well as possible links between the distribution of body fat and cardiovascular diseases. The system consisted of nine CCD cameras arranged in three stereo triples, a 386-based personal computer equipped with two frame grabbers, and three slide projectors for projecting dot patterns on the body. A method was developed for the automatic and simultaneous calibration of the interior and exterior orientations of all cameras. Algorithms were developed for the automatic measurement of image coordinates, for finding corresponding match points in a stereo triple, integration of multiple stereo models, and for computation of volumes, surface areas, and circumferences. About 15 minutes of time was required to capture 18 stereo triples of a human subject. Preliminary test achieved an accuracy of better than 5% in measuring body volume.

KEY WORDS: Computer vision, mapping, human, bodies, automation.

1. INTRODUCTION

Measurement of human body dimensions and volume has always been an important task in the study of growth pattern, body composition, physical health, and motor skill development. After over thirty years of continuing research in these areas, the Department of Kinesiology in the University of Illinois at Urbana-Champaign (UIUC) has accumulated a large quantity of data based on such measurements as height, weight, skinfolds, circumferences, and volume from underwater weighing. Recent research has indicated that much more detailed measurements about the human body are desirable. For example, body surface area is known to be directly related to body composition and energy expenditure, yet the commonly used method is to compute the surface area using height and weight in a formula developed in 1916 (DuBois and Dubois, 1916). Recent studies have also shown that the distribution, as well as total amount, of excess body fat have direct bearing on the risk of cardiovascular disease (Buchard and Depres, 1988). It is obvious that advanced research into these problems can benefit significantly from the development of an automated, accurate, inexpensive, safe, and unobtrusive method of mapping the entire human body.

The use of stereo photography to measure the size and shape of the human body was investigated as early as 1916 (Benedict, 1916). In 1967, through a cooperative project between the Department of Civil Engineering and the Children's Research Center at UIUC, Weissman and Herron developed a method to use stereo photogrammetry to accurately measure the volume and surface area of handicapped children (Weissman and Herron, 1967). Subsequent works by Herron, Karara and others led to the development of a completely new technology, called biostereometrics, which deals with the measurement of human body parts (Karara, 1989). Conventional methods of stereo photogrammetry require expensive equipment such as metric cameras, stereoplotters and highly-trained operators. Moreover, it is extremely tedious and time consuming to measure the thousands of points needed to represent a human body. As a result, these methods are not in common use at present.

The solution lies in the complete automation of the method of stereo photogrammetry. Recent developments in solid-state cameras and image processing techniques have resulted in a new technology called computer vision and have created the opportunity for performing automated, metric measurements using low-cost cameras and personal computers. Recent research have clearly demonstrated the potential capability of computer vision for performing accurate geometric measurements in an automated mode (El-Hakim, 1986; Wong 1986; Maas, 1991). The mapping of an entire human body using computer vision techniques presents special problems. The smooth texture of human skin provides little contrast for image matching, and dictates the need for a projected light pattern. In order to overcome the problems of 1) occlusions in mapping a three-dimensional body, and 2) small dimensions (about 8.5x6.4 mm) of the focal planes of vision cameras, a large number of stereo views from many different view angles must be acquired. As the number of stereo views increases, the complexity of data reduction also increases.

This paper describes a highly automated computer vision system that has been developed to map the human body. The system was assembled from commonly available components, and all the measurement tasks are performed by software that can be run on either personal computers or workstations. The system was designed so that it can be easily adopted for the mapping of any objects, such as machine parts in industrial gaging and architectural details in engineering construction.

2. IMAGING SYSTEM

The vision system was assembled from components that were available in the Vision Laboratory of the U.S. Army Advanced Construction Technology Center at UIUC. A Zenith 16-Mhz 386/AT personal computer, with 4MB RAM, an 80-MB hard disk drive, one 5.25-inch and one 3.5-inch high density drives, was used as the control computer. Two EPIX frame grabbers, each with a 20-Mhz pixel clock and 1 MB of image memory, were installed in the PC. An EPIX interactive program called Silicon Video MUX was used to digitize, process, display, transmit, and archive video information (EPIX, 1989). Video data was digitized in 256 grey levels with a resolution of 480x1040 picture elements (pixels). Each frame grabber was equipped with an internal switch which could select video input from up to 6 cameras under computer control. Nine charge-coupled-device (CCD) cameras were connected to the frame grabbers. These included the following:

4 Pulnix TM-80 frame-transfer CCD cameras, 2/3-inch focal plane, 490x800 pixels, Fujinon 12.5-75mm, F1.2 zoom lens

2 Pulnix TM-845 frame-transfer CCD cameras, 2/3-inch focal plane, 490x800 pixels, asynchronous variable shutter (1/8000 sec to 1/60 sec) Fujinon 12.5-75mm F1.2 zoom lens;

3 General TCZ-200 interline-transfer CCD cameras, 2/3-inch focal plane, 492x510 pixels, Computar 12.5-75mm F1.8 zoom lens.

The six Pulnix cameras were essentially identical in characteristics. The electronic shutters of the TM-845 cameras were not utilized during image capture. The six Pulnix cameras all have a resolution of 525

TV lines along the horizontal direction (H) of the scan lines, and 350 TV lines vertical (V) to the scan lines. The three General cameras have a lower resolution of 370(H)x 350(V) TV lines.

The nine cameras were grouped into three units of triple, with each triple consisting of two Pulnix cameras and one General camera as shown in Figure 1.

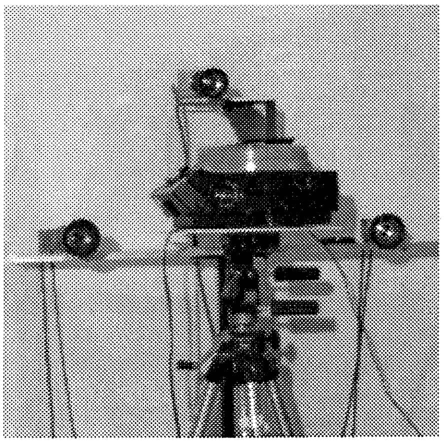


Figure 1. Stereo triple of three cameras with a slide projector

The two Pulnix cameras were mounted on a horizontal bar, separated by a base distance of about 73 cm, and with the two optical axes converging at an angle of about 15 degrees. The General camera was mounted at mid-point between the two Pulnix cameras, but elevated about 30 cm about the horizontal bar supporting the two Pulnix cameras and tilted downward by about 10 degrees. All three cameras were rotated 90° about the optical axes, so that the scan lines are along the vertical direction. Studies have shown that geometric accuracy is slightly higher perpendicular to the direction of scan than along the scan lines. By rotating the cameras, the direction of higher geometric accuracy was used for computing the distance of object points from the cameras by spatial intersection.

Positioned immediately in front the cameras was a Kodak Carousel 4400 slide projector, equipped with a Kodak Ektanar C 102-152mm, f/3.5 zoom lens. The projector was used to project a pattern of black dots onto the body surface of the subject during image capture.

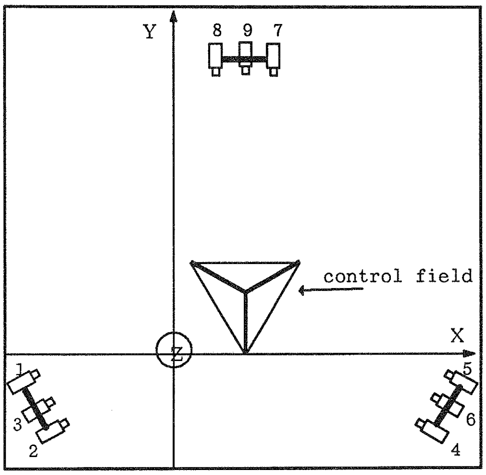


Figure 2. Laboratory setup for image capture

Figure 2 shows the arrangement of the three stereo triples during image capture. The arrangement was designed to provided stereo coverage of a space measuring about 1mx1mx1m, with the center located about 1 m above the floor and about 1.8 m from the cameras. The resulting object/image scale was about 2 mm/pixel.

3. IMAGE CAPTURE

In order to obtain complete stereoscopic coverage of an entire body, six separate poses were needed for each human subject:

Pose 1: head and upper body;
Pose 2: lower body trunk;
Pose 3: left leg;
Pose 4: right leg;
Pose 5: left hand; and
Pose 6: right hand.

During each pose, images were captured in the following sequence:

1. Turn projector 1 on, and capture images with cameras 1, 2, and 3 in sequence. Turn projector 1 off.
2. Turn projector 2 on, and capture images with cameras 4, 5, and 6 in sequence. Turn projector 2 off.
3. Turn projector 3 on, and capture images with cameras 7, 8, and 9 in sequence. Turn projector 3 off.

The subject was asked to remain still during the entire image capture procedure of each pose, which could be completed in about 1 minute of time. The cameras remained fixed in place during the entire imaging session, while the subject must be moved between poses to present different section of the body to the cameras for imaging. About 15 minutes of time was required to capture all the images for one subject, consisting of 18 stereo triples and 54 images. Figure 3 shows the three images of one pose from cameras 1, 2, and 3.

The nine images of each pose were used to generate a digital model of a section of the body. Thus, the six poses actually resulted in six digital models of different sections of the body. To provide tie points between the six different sections, 1/4-inch diameter round retroreflective targets were placed on the body. These targets appeared as white dots in the images shown in Figure 3.

4. CAMERA CALIBRATION AND ORIENTATION

A portable three-dimensional control field, see Figure 4, was built to perform camera calibration. It consisted of three 6.33mm-thick aluminum panels mounted vertically at an angle of 120-degrees with respect to each other. The panels were painted white. On each panel were thirty-three 0.8mm-diameter holes positioned to an accuracy of ± 0.03 mm with respect to each other. Surrounding each hole is painted a black 22.86mm-diameter round target, which was identified with an eight-digit bar code. There were 198 targets in the control field. The three-dimensional coordinates (X, Y, and Z) of the targets were computed from their spacing within each panel and from the orientations of the three panels. It was estimated, and later verified by camera calibration results, that these coordinates had an estimated root-mean-square (RMS) error better than ± 0.1 mm.

Before each imaging session, the portable control field was placed in the area to be occupied by the subject, as shown in Figure 2. An image of the control field was acquired from each of the 9 cameras. A computer software package was developed to perform the following computations:

1. determine the image coordinates of the centers of all targets identified in the image;
2. identify the code number of each of the target;

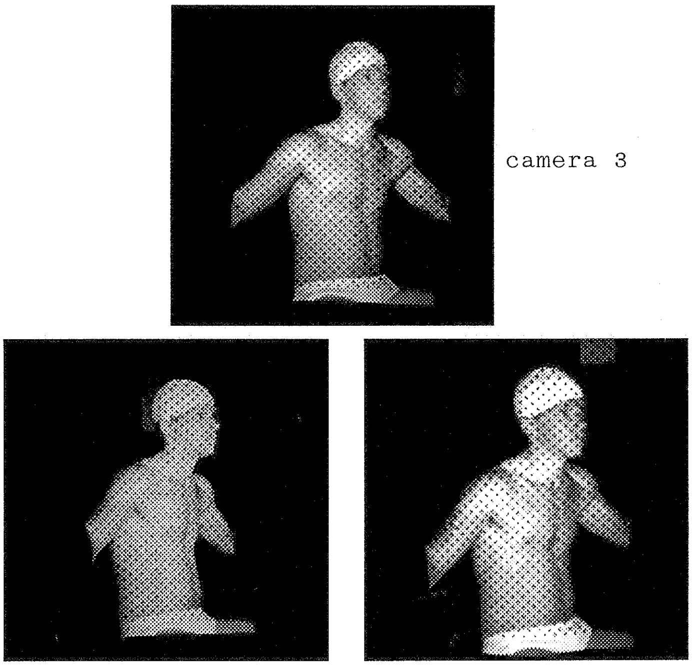
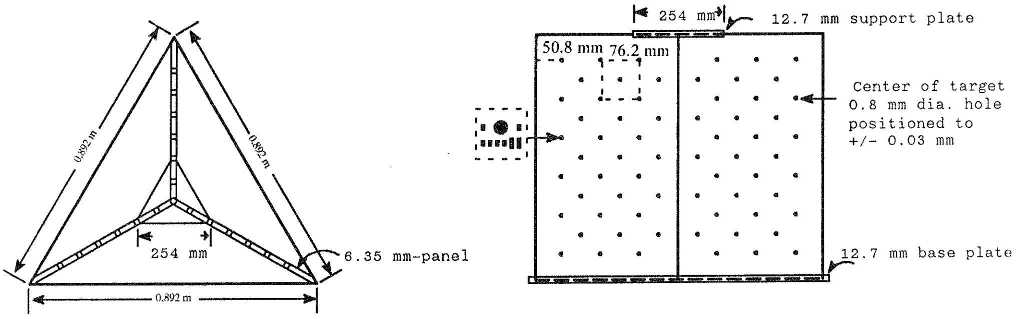


Figure 3. Image from a stereo triple



(a) Top view (b) Side view

Figure 4. A Portable control field

3. compute preliminary estimates of the camera's exterior orientation parameters by the method of direct linear transformation (DLT); and
4. computer the interior and exterior orientation parameters of the cameras by a bundle adjustment.

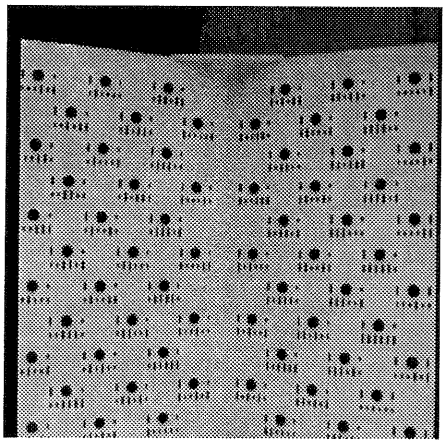


Figure 5. Image of control field from a camera

Figure 5 shows an actual image used for the calibration of one camera. The distortion model is represented by the following expressions:

$$dx = \bar{x}(l_1 r^2 + l_2 r^4) + [p_1(r^2 + 2\bar{x}^2) + 2p_2\overline{xy}]\,[1 + p_3 r^2]$$

$$dy = \bar{y}(l_1 r^2 + l_2 r^4) + [2p_1\overline{xy} + p_2(r^2 + 2\bar{y}^2)]\,[1 + p_3 r^2]$$

$$\bar{x} = (x - x_p)(1+k)$$

$$\bar{y} = y - y_p$$

$$r^2 = \bar{x}^2 + \bar{y}^2$$

where x and y are image coordinates; x_p and y_p are image coordinates of principal point; k is a scale correction factor for the x coordinates; l_1 and l_2 are terms for radial lens distortions; and p_1 and p_2, and p_3 are the terms for asymmetric lens distortions. It was found that the l_2 and p_3 terms could be set equal to zero for all the CCD cameras used in this project.

Table 1 shows the calibration results for three cameras within a triple arrangement; where the object-space coordinates of the exposure centers of the cameras are given by the coordinates (X^c, Y^c, Z^c) and the orientation angles are represented by ω, φ, and κ). Typically, the standard errors of the image coordinate residuals (σ_x, σ_y) ranged from ± 0.05 to 0.10 pixel. The variation in image residuals was partially caused by the difference in orientations of the cameras with respect to the control field. A camera that was pointed directly at the intersection of two control panels would result in a stronger resection geometry than a camera that was pointed more to one side. Moreover, greater tilt with respect to the control panels also resulted in greater perspective distortion of the shape of the circular targets and therefor lower accuracy in finding the centers of the targets in the image. It is interesting to note in Figure 6 that, in spite of the lower resolution capability of camera No. 3 (General), it actually had a smaller image residuals than camera No. 2 (Pulnix) in this particular case. It was most likely caused by the slightly poorer imaging geometry of camera 2 with respect to the control field.

5. DATA PROCESSING

One of the major advantages of using the computer vision approach for mapping is the ability to compute object-space coordinates directly from the images without making any manual measurements. A software package was developed to compute object-space coordinates of points on the body by the following procedure:

1. for each image, compute the image coordinates of the centers of all of the projected points on the body;
2. for the three images within each triple, find the matching points within the three images from the constraints of epipolar geometry (Mass, 1991);
3. compute the object-space coordinates of the points imaged in each triple by spatial intersection, using results from camera calibration; and

Table 1. Sample results from camera calibration

Parameters	Camera 1 (Pulnix)	Camera 2 (Pulnix)	Camera 3 (General)
x_p	4 ± 2 pixels	6 ± 3 pixels	23 ± 3 pixels
y_p	-26 ± 3 pixels	-12 ± 2 pixels	-2 ± 4 pixels
f	985 ± 3 pixels	973 ± 3 pixels	1065 ± 5 pixels
k	-0.0218 ± 0.0001	-0.0205 ± 0.0003	-0.04178 ± 0.0003
L_1	0.293E-6 ± 0.5E-8	0.29E-6 ± 0.6E-8	0.992E-7 ± 0.7E-8
P_1	-0.110E-5 ± 0.3E-6	-0.44E-6 ± 0.6E-6	-0.13E-5 ± 0.8E-6
P_2	0.523E-6 ± 0.3E-6	0.97E-6 ± 0.4E-6	-0.45E-5 ± 0.8E-6
X^c	-121.9 ± 0.5 cm	-70.0 ± 0.4 cm	-96.3 ± 0.6 cm
Y^c	-53.9 ± 0.3 cm	-103.3 ± 0.5 cm	-80.7 ± 0.6 cm
Z^c	70.2 ± 0.1 cm	70.20 ± 0.06 cm	101.0 ± 0.2 cm
ω	91.1 ± 0.3 deg	91.8 ± 0.2 deg	79.6 ± 0.3 deg
ϕ	-59.2 ± 0.1 deg	-36.7 ± 0.2 deg	-48.1 ± 0.1 deg
κ	1.4 ± 0.2 deg	1.13 ± 0.1 deg	-8.2 ± 0.2 deg
σ_x	± 0.06 pixel	± 0.1 pixel	± 0.09 pixel
σ_y	± 0.08 pixel	± 0.1 pixel	± 0.08 pixel

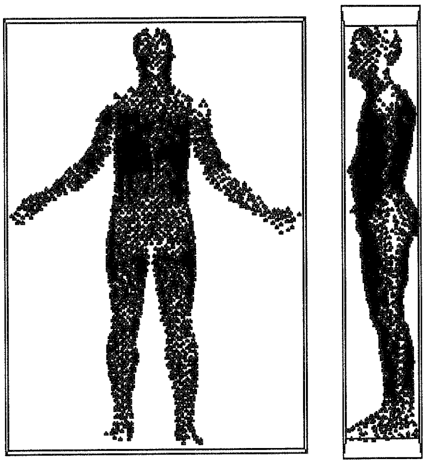
4. merge the object-space coordinates from the six poses by performing three-dimensional coordinate transformation using the tie points.

A software package was developed to generate cross-sections of the body, and to compute the body volume and surface area.

All the data processing were performed on an UNIX-based Apollo workstation for faster computing speed and larger storage capability. Transfer of image files from the 386-PC was accomplished with floppy disks.

6. RESULTS

Figure 6 shows two views generated from the 3,400 points that were computed to represent one subject. Figure 7a shows a cross-section of the same subject generated from the digital model, and Figure 7b shows the same cross-section after data smoothing and densification. Figures 8 and 9 show the volume and surface area distributions respectively. The total body volume was computed to be 73,897 cm^3, as compared to 70,530 cm^3 measured by the dunk tank method. In the dunk tank method, the subject was weighted first out of water, and then again when fully submerged in water. While being weighed in water, the subject was carried in a basket in a crouched position. The water density was also measured. The body volume was then computed from the measured weights and water density. Considering the difference in postures used in the two methods, the agreement in the computed volumes indicated the potential accuracy of the computer vision approach. Further testing of the method is being conducted with additional subjects.



a. Front view b. Side view

Figure 6. Digital model of a subject

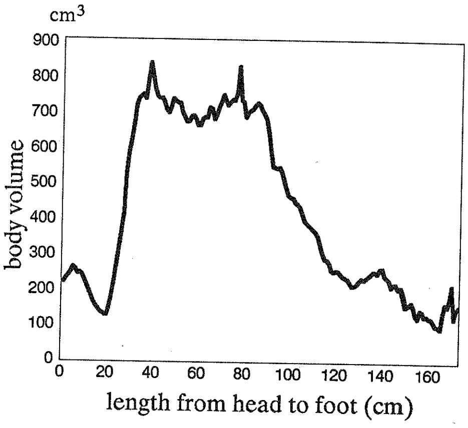


Figure 8. Body volume distribution

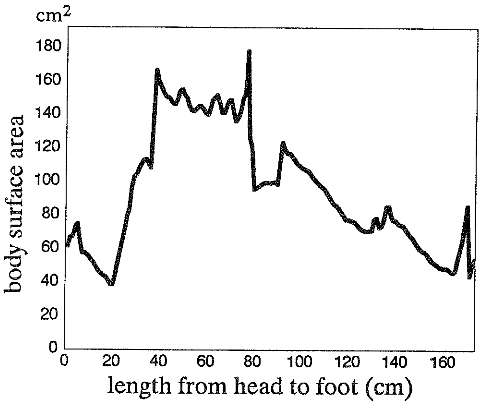
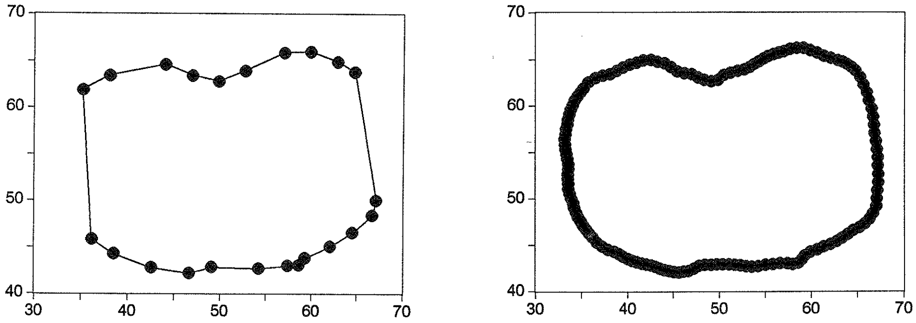


Figure 9. Body surface area distribution



a. from digital model b. After smoothing and densification

Figure 7. Generated cross-section

7. CONCLUSIONS

This project clearly demonstrated the potential capability of vision system in biostereometric applications. The major advantage of computer vision approach is the ability to completely automat the entire data reduction process and to produce results in matters of hours, rather than days. In this demonstration project, the only step that required manual processing was the matching of tie points between different sections of the bodies. This step can also be automated in the future by using different shapes of targets for tie points, or by using targets identified with a bar code.

The design of the vision system presented in this paper was constrained by the equipments that were available in our laboratory. The three triple arrangement shown in Figure 2 did result in the lack of stereo coverage on some parts of the body. With the addition of three more cameras to provide a fourth triple, the system should produce results of even higher accuracy in measuring body volumes and surface areas.

The human body is one of the most difficult object to measure accurately, because of its lack of surface texture, its shape, its inability to remain still for even a few minutes of time, and the inherent necessity to accommodate the emotions and feelings of the subject during imaging. The methodology reported in this paper can be easily extended for the mapping of other objects, such as machine parts, architectural details in buildings, and clay models used in reverse engineering.

8. ACKNOWLEDGEMENT

The research reported in this paper was conducted at the University of Illinois at Urbana-Champaign and supported in part by the University Research Board, the Manufacturing Research Center, and the Advanced Construction Technology Center which was being supported by the U.S. Army Research Office under the DoD-University Research Initiative Program.

9. REFERENCES

Benedict, F. G. "A Photographic Method for Measuring the Surface Area of The Human Body." Am. J. of Physiol., 41:275-291, 1916.

Buchard, C. , and J. Depres. "Variation in Fat Distribution with Age and Health Implications." Physical Activity and Aging. The Academy Papers, The American Academy of Physical Education No. 22, 1988.

Dubois, D., and E. F. Dubois. "A Formula to Estimate the Approximate Surface Area if Height and Weight be Known." Arch. Intern. Med., 17:863-871, 1916.

El-Hakim, S.F., 1986. Real-time image metrology using CCD cameras. PE&RS, 52(11), 1757-1766.

Herron, R.E., 1972. Biostereometric measurement of body form. Yearbook of Physical Anthropology, Vol. 16, 80-121.

Karara, H. M. (Editor-in-Chief). Non-Topographic Photogrammetry. Second Ed. American Society of Photogrammetry and Remote Sensing, 1989.

Mass, Hans-Gerd, 1991. Digital photogrammetry for determination of tracer particle coordinates in turbulent flow research. PE&RS, 57(12), 1593-1597.

Weissman, S. and R. E. Herron. "Stereophotogrammetry as a Means of Anthropometry for Mentally Handicapped Children." A report on a study sponsored by Illinois State Department of Mental Health and National Institute of Mental Health. Civil Engineering Studies, Photogrammetry Series No. 11, University of Illinois, November 1967.

Wong,K. W. and Wei-Hsin Ho. "Close-Range Mapping with a Solid State Camera," Photogrammetric Engineering and Remote Sensing, Vol. LII, No. 1, pp. 67-74, 1986.

Wong, K. W.: A. G. Wiley; and M. Lew. "GPS-Guided Vision Systems for Real-Time Surveying." Journal of Surveying Engineering, Vol. 115, No. 2, pp.243-251, 1989.