INVESTIGATION OF GENERATING 3D POINT-CROWD FOR HUMAN MODELLING USING HYBRID VIDEO THEODOLITE

T. Anai *, H. Chikatsu

Tokyo Denki University, Dept. of Civil and Environmental Engineering,
Hatoyam, Saitama, 350-0394, JAPAN
(anai, chikatsu)@g.dendai.ac.jp

Commission V, WG III

KEY WORDS: Analysis, Extraction, Image, Sequence, Dynamic, Modelling

ABSTRACT:
Generally, dynamic analysis of human motion using video image sequences is performed with some markers fitted on human's feature points. Therefore, 3D coordinates for human feature points are obtained in real-time. However, these markers should be removed because these markers often obstruct natural human motion. In order to remove the markers on human body, though 3D human modelling is one of the effective methods, some problem should be resolved. These problems include, automatic camera calibration, human extraction from complex background image, automatic stereo matching and or so on.
On the other hand, the authors have so far developed a Hybrid Video Theodolite system for multiple applications, and dynamic analysis of human motion using the HVT system has been performed. With these circumstances, the authors investigate an effective method of generating 3D point crowd for human body using image sequence acquired the HVT System in this paper.

1. INTRODUCTION

Recently, 3D measurement using video image sequences is widely performed in the field of dynamic analysis of human motion with appearance of functional image processing equipment. Generally, dynamic analysis of human motion has been performed under artificial condition such as camera position and rotation are fixed, and background and lighting environment are controlled. Moreover, some markers are fitted on the human feature points. Therefore, it is possible to calibrate camera parameter previously, and 3D coordinates of human feature points are computed in real-time. However, in order to analyze the most natural human motion, limitations for camera and test field should be removed. Also, the markers on the human body should be removed.
In order to perform gait analysis with ideal condition, the authors have been concentrated on developing the Hybrid Video Theodolite (HVT) System. Figure1 shows the appearance of Hybrid Video Theodolite system. The HVT system is a multi system which consists of two color CCD cameras for stereo image acquisition, one CCD camera for tracking a moving object, motor driven PAN head and TILT body, laser range finder for camera calibration, time generator, synchronizer, personal computer and VTR. The remarkable feature of this system is its ability to synchronize stereo image sequences and rotation parameters while tracking a moving object. As for further additional feature of this system, automated camera calibration without target has been achieved.
Camera calibration for the HVT system have been investigated (Chikatsu & Anai, 2000) and dynamic analysis of human gait motion with constant background test field using markers on human feature points has been demonstrated (Anai & Chikatsu, 2000).
On the other hand, in order to perform marker-less dynamic analysis of human motion, 3D modelling of human body becomes important issue. For this goal, the authors investigated an effective method of generating 3D point crowd data using stereo image sequences acquired by the HVT system.

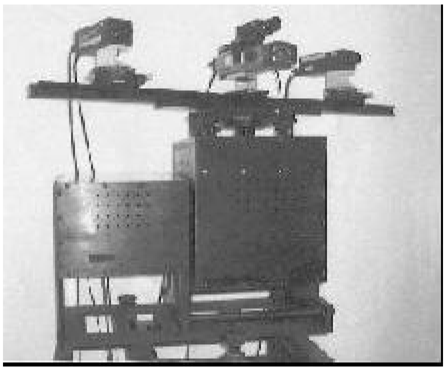

Figure 1. Hybrid Video Theodolite system

2. HUMAN GAIT MOTION EXPERIMENT

In order to evaluate the effectiveness of the HVT system for generating 3D point-crowd of human body, the experiment of human gait motion was performed. In this experiment the subject walked about 10m in the straight line, and the distance from the HVT system to walking course was about 5m.
Camera calibration for the stereo cameras was performed by automatic camera calibration technique using the HVT system.
Figure 2 shows the flow of generating 3D point-crowd of human body in this experiment. Details are as follows..

* Corresponding author.

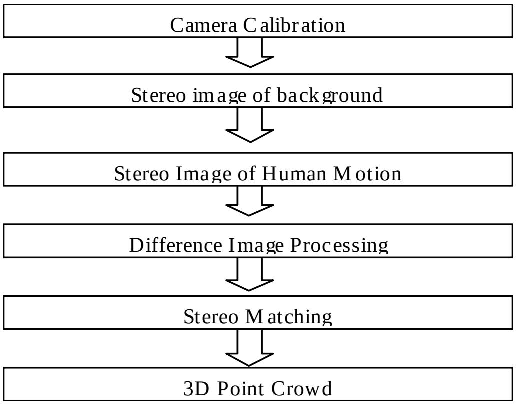

Figure 2. Flow of Generating 3D Point-crowd

2.1 Stereo image of Background

Human extraction is one of the important issues in the field of dynamic analysis of human motion. Generally, difference image is known as the efficient methods for human extraction. In this experiment, in order to perform the human extraction using difference image, stereo image sequences for background of test field were obtained in advance.
By the way, the HVT system has function to generate time at interval of 1/100 second and times for each image are continuously superimposed on image frames. Furthermore, time information is synchronized to the rotation parameter of PAN head and TILT body. Therefore, difference images are obtained automatically by applying this function to background and foreground image.
Figure 3 shows concept for generating difference images using the HVT system.

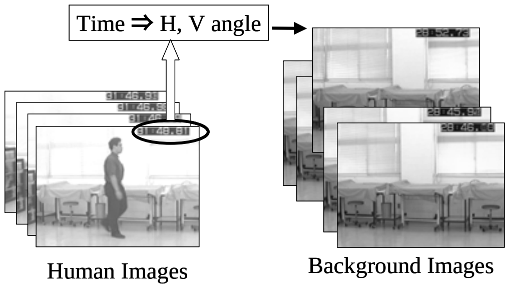

Figure 3. Concepts for generating difference images

Figure 4 shows the stereo images of background in this experiment. The upper right-hand number is superimposed time by the HVT system, and it can be understood that the stereo image sequences are synchronized.

(a) Left Image (b) Right Image

Figure 4. Stereo Image for Background

(a) Left Image (b) Right Image

Figure 5. Stereo Image of Human Gait Motion

2.2 Stereo Image of Human Motion

After the acquisition of background image, stereo image sequences of human gait motion were acquired. Figure 4 shows the stereo image of human walking.
In this experiment, the subject's clothes were not limited. Moreover, information on the rotation parameter of the HVT is associated with each frame of the stereo image sequences as well as the background image.

2.3 Difference Image Processing

As mentioned above, the background image has been decided automatically. However, the background of test field and lighting environment condition were not constant in this experiment. Therefore, in order to remove the influence of shadow, lighting or noise, difference image processing were performed as follows.

- The average value and the standard deviation around interest point of background are computed using 31 ×31 mask.
- Similarly, the average value around interest point of foreground image is computed.
- The difference value of interest point between foreground image and background image is computed using average value of each image.
- If the difference value of interest point is larger than standard deviation of background image, interest point is estimated as human area.

Figure 6 shows the result of human extraction, and it can be seen that human area was extracted completely. However, background area around human body still remains. These background areas are able to remove in the stereo matching procedure.

2.4 Stereo Matching

2.4.1 Generating Point-Crowd on the Human Area

In order to generate 3D point-crowd of human body, SUSAN operator was performed for the left image as the first step (Smith, S. M. and Brady, 1997). Figure 7 shows generated point crowd.

2.4.2 Least Square Matching

Matching for the generated point-crowd was performed by Least Square Matching (LSM) (Grün, 1985). The LSM is known as the robust methods of area matching. The remarkable point of the LSM is that the gray level differences between the template and the matching window is minimized with changing the position and the shape of the matching window.

On the other hand, epipolar lines is one of the important information for stereo matching, and epipolar line is computed using calibration parameters obtained by the HVT system. Therefore, the LSM was performed efficiently along the epipolar line.

Moreover, in order to remove miss-matching points, back matching was performed in this paper. Figure 8 shows the result of the stereo matching.

(a) Left Image (b) Right Image

Figure 6. Result of difference image processing

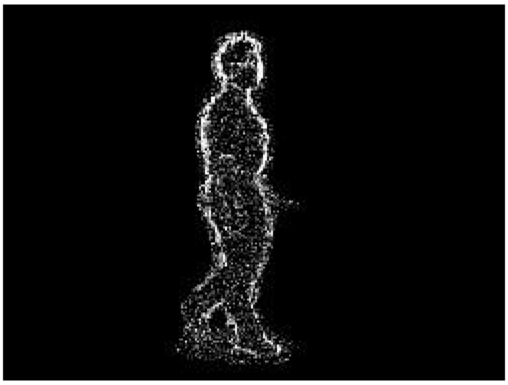

Figure 7. Result of SUSAN operator

(a) Left Image (b) Right Image

Figure 8. Stereo matching

Figure 9. 3D point-crowd

3. 3D POINT-CROWD

3D point-crowd for human body is generated as the results of stereo matching. However, the points of the background or miss matching still remain. By the way, the distance from the HVT system to walking course was about 5 m, and the subject walked on the walking course. Therefore, filtering for 3D point-crowd using the depth information was performed, and the depth value between 4.8m and 5.4m was estimated as human body.

Figure 9 shows the 3D point-crowd, which was filtered by the depth information. From this image, it is concluded that the generating 3D point-crowd using the HVT system is effective method for the human modelling or marker-less dynamic analysis of human gait motion.

4. CONCLUSION

In order to perform marker-less dynamic analysis of human motion, the authors developed the effective method for generating the 3D point-crowd of human body using function of the HVT system in this paper.

In this method, the automatic human extraction using difference image processing was performed and effectiveness of the HVT system for human extraction is descried. Similarly, automatic stereo matching for human body using the LSM and epipolar line were investigated. Finally, generating the 3D point-crowd from stereo image was performed.

However, there are still some problems, which need to be resolved, before this system becomes operational. These problems include that the more effective human extraction from video image sequences, stereo matching for the low information area on the human body, and increasing the speed of these procedure.

5. REFERENCES

Chikatsu H. and Anai T, 2000, Development of Hybrid Video Theodolite and Application for 3D Modelling of Panoramic Objects, International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B5, pp130-pp136, Amsterdam.

Anai T. and Chikatsu H., 2000, Dynamic Analysis of Human Motion using Hybrid Video Theodolite, International Archives of Photogrammetry and Remote Sensing, Vol. XXXIII, Part B5, pp25-pp29, Amsterdam.

Smith, S. M. and Brady, J. M, 1997, SUSAN-A New Approach to Low Level Image Processing, International Journal of Computer Vision, Vol.23, No.1, pp.45-78

Grün, A., 1985 Adaptive Least Squares Correlation: A Powerful Image Matching Technique, South Africa Journal of Photogrammetry, Remote Sensing and Cartography, Vol.14, No.3, pp.175-187