Stereo Camera based Markerless Motion Capture by Morphological Analisys of Disparity Map

Sewoong Jun
Intelligent Robotics Research Center
KETI
Bucheon, Korea
daniel@keti.re.kr

Bongseok Kim
Intelligent Robotics Research Center
KETI
Bucheon, Korea
Seize95@keti.re.kr

Abstract—**This paper presents a new real-time system to acquire motion information of human articulated objects such as arm and head. The system does not need any marker or device to wear on human body and adopted stereo camera to obtain robust system against for illumination and complex background without position initialization of articulated objects. We present a solution to estimate self-occluded body objects when human model behaves normal action towards the camera. The main idea of the solution is to apply a component labeling techniques on sliced disparity map, and found the arm position when the arm is located in front of basis distance of body and we could also found arm location when the arm is located on the basis distance with Morphological methods. From this approach, we can obtain the full body shape considering self-occlusion. It is simple and fast in comparison with other methods which satisfy real-time performance and accuracy of object tracking at the same time. As the result of experiments, we obtained the reliable motion data of articulated object motion under various costume and body type of models.**

Keywords-component; HRI, Stereo Camera, Markerlessmotion capture, Disparity Map

I. Introduction

Recently the ubiquitous computing which can communicate with the human in anywhere without the user's recognition of the system, has been noticed as the new paradigm of computerization. In the ubiquitous environments, the computer needs to obtain user's intention or information with accuracy. For this reason, the Human Computer Interaction (HCI) has occupied an important position.

Especially, Computer Vision base HCI system has unique advantage which is non-tangency and non-awareness of sensor's existence.

In vision-based HCI, there are many perception modalities. These can be divided into six categories such as [2]: people tracking, gesture recognition, facial perception, facial expression, eye orientation tracking and head pose tracking. Especially, Human motion information that combines people tracking and gesture recognition has a unique advantage which is the simple and general approach to obtain user's intention. Information of motion is an intuitive expression cue for estimating user's intention. We can simply obtain user's intention by estimating the human motion information. And human motion information can be obtained analyzing continuous location of articulate body part of human.

Proposed method can be used to ubiquitous system or robot system. When the system understands human needs, they can perform what users want. Moreover, the method can be use in base research skill for human ethology.

Computer vision based motion capture system is can be divided into two sorts of system extensively. First one is Marker based human motion capture system and other one is Markerless motion capture system. Marker based human motion capture system is widely used on computer 3D-graphics and there are many commercial products. The user can obtain accurate result, but user needs to learn how to use it and many pre-calibrating for different body types for targets. On the other hands, Markerless human motion capture system does not need any device which the user wears equipment for the system.

Many of previous researches were proposed on this topic. Leung[3], Kakadiaris[4] and Lerasle[5], have developed an ribbon model, deformable model, full body model for their algorithm. However, they can't cover self-occlusion and complex pose such as the motion that hand on user's head. Recently, there are many efforts on this research field to overcome above problems. The method of using stereo camera and multi camera were proposed and using silhouette method was also suggested.

Our method also based on stereo camera. However, we proposed novel algorithm that can estimate self-occlusion on real-time performance. We modify disparity-map from stereo camera and we make it blob with suggested filter. And then, the blob was separated by occlude articulated object section and non-occlude articulated object section. The algorithm is applying different method on each section that we suggested. The each result could merge and we could obtain reasonable results.

This paper is organized as follows: In section 2, our algorithm is presented. First, we describe preprocessing procedure before our algorithm work. Second, the main feature is introduced and explained how to acquire. Then third, we explain human motion capture algorithm in case of self-occlusion and finally, Morphological approach is described in case of without self-occlusion. In section 3, we cover the system setup with detail specification. In section 4, we show the experimental result of our system. And finally in section 5, we evaluate our algorithm and conclude this paper.

II. MARKERLESS HUMAN MOTION CAPTURE

The procedure of the system has divided into 4 steps approximately. Every procedure has been performed in computer software on system. The output of the algorithm is left hand, right hand, body center and head position in 3D working space.

A. Preprocessing

To obtain precision disparity map, we need a fine camera calibration. An accurate intrinsic and extrinsic camera parameter is essential for the disparity matching procedure. In this paper, we used the commercialized library, which is Smallvcal and VidereDesign's camera calibration program [7,9].

B. Main Feature Extraction from Disparity Map

From the calibration of stereo camera we could obtain accurate disparity map, Videre-Desigh STH-2 Stereo camera can suggest 3D map information within 1~2cm error rate on about 2M distance in front of stereo camera. And then, We adopt noise blob removal filter on the disparity-map and find body center location by COG(Center of Gravity) consequently. Fig 2. shows overall follow of Main Feature extraction method from disparity map.

First, we use Noise blob removal filter by counting continuous disparity pixels. Count number of pixels mean size of disparity blob and the filter remove 1 or 2 pixel size blob immediately.

Second, we remove background. We assume that there are only one person on the working space and target human stand-up perpendicularly. From this, we could obtain Center of Gravity (COG) of the all disparity map throughout Noise blob removal filter procedure. Then, we remove all disparity map 30cm behind of COG.

Next, we could estimate COG of Human Body, this is different point compare with COG of all disparity map. We find first left and right disparity blob by searching from left and right side of bottom of disparity image and find body area of disparity (red area of Fig. 3).

We assume that only one target human is located in front of camera and stand up toward the cameras. The system can analyze the human body disparity from back ground by histogram of disparity map. The histogram present enough cue to distinguish human body and we could obtain human distance from camera by measuring disparity of body COG(Center of Gravity).

We also could find hand blob when the hand part is located in front of body distance from the stereo camera. Fig .4 shows

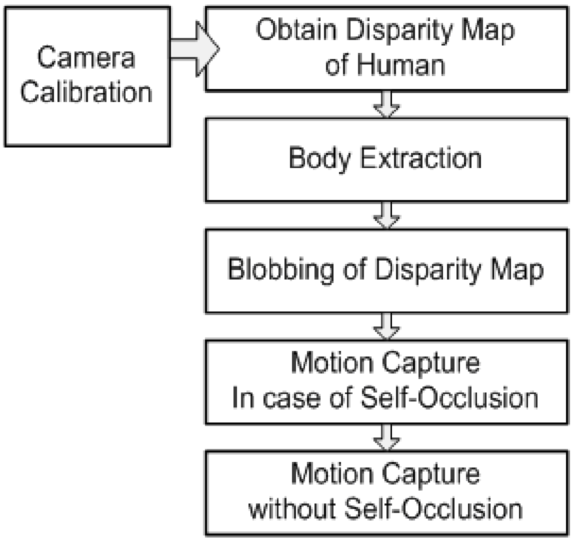

Figure 1. The overall procedure of the system

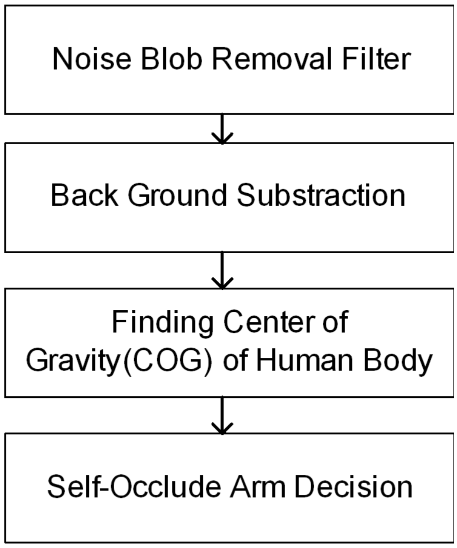

Figure 2. Main feature extraction procedure

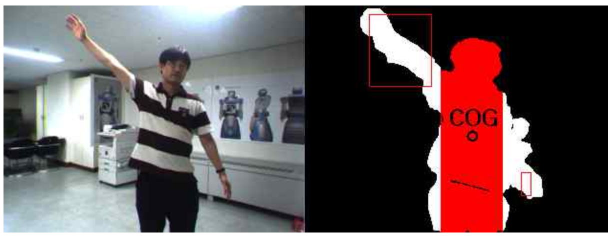

Figure 3. Finding center of gravity (COG) of target human body.

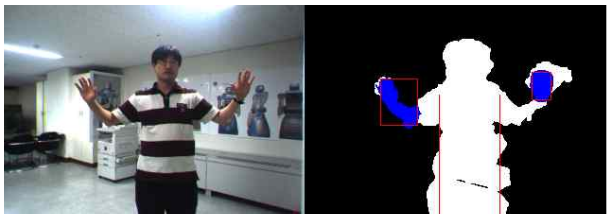

Figure 4. Finding hand blob when the hand is located in front of body distance from the camera.

Identify applicable sponsor/s here. If no sponsors, delete this text box. (sponsors)

hand blob parts. Hand blob part number could be 0, 1, or 2. This method was adopted by previous research [10] using component labeling and self-occlusion blob extraction algorithm.

Finally, we could defined the still human motion disparity map image has occlude arm blob or not.

C. Human Motion in case of Self-Occlusion

From the previous procedure, the system extracts blobbing area which assumes that occlude parts of arm in front of human body in distance from the stereo camera.

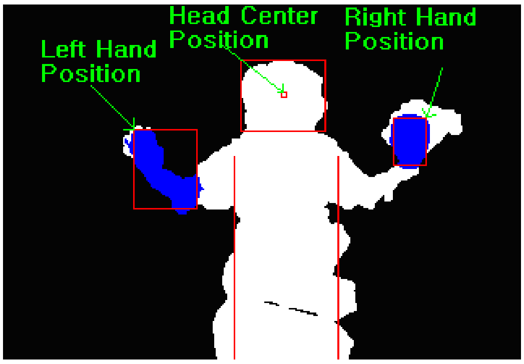

Figure 5. Hand Position Estimating in case of self-occlusion

In this case, each hand position is estimated by edge point of blob area which is red rectangular on Fig. 5. And we distinguish left and right hand from the point of opposite diagonal corner point of left and right hand positions by comparing with x location of disparity map image.

We already know human body center position from previous procedure. We also know the body distance from the camera and body size (red part on Fig. 3). From this, we could estimate head center position on disparity map by following equation (1) and (2).

$$Head_{(x,y)} = \begin{cases} X = COG_x \\ Y = (COG_y - L_y) \end{cases} \tag{1}$$

$$L_y = \frac{C \cdot P_n}{D_z} \tag{2}$$

where X and Y is head position of disparity map and COGx is center position of target body, Pn is pixel number of body part of disparity map, Dz is body COG distance from the stereo camera and C is camera and target dependent constant value by obtaining from experimental trial and errors.

Finally, we obtain 4 positions from disparity map image and we could replace them easily to 3D location in the working space by the function of SmallV[9] of Videre-Design.

D. Human Motion without Self-Occlusion

We could find up easily on previous section when the self-occlusion was detected. However, many undetectable pose are still remained such as hands up pose and attention pose. In this case, hand is located on same distance of body COG. Thus, we need to use different method to find hand location.

The presented algorithm finds one side of articulated object of human by morphological approach. It can be hand or elbow. We evaluate shape of human body and decide that it is hand or elbow. Finally we could obtain human hand geometry information considering hand like blob which explained section 3.2.

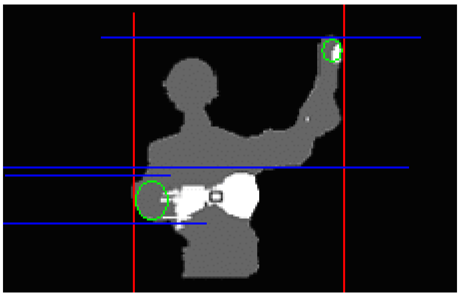

Figure 6. Morphological human motion capture in case of non self-occluded human body

The algorithm compares 3 different morphological states from disparity map in case of non self-occlusion. Each case detects hand position. It shows on Fig. 7 and the algorithm choices one position by measuring distance from head and COG of body.

Finally we could obtain head center position and body center position, left hand and right hand position in case of non self-occlusion case.

III. Experimental Result

We implement our algorithm using C++ under the Pentium 4 PC with Microsoft Windows environment and the Videre Design MEGA-D stereo camera is connected by IEEE1394 cable. We use the Videre Design SVS library to estimate the disparity map on real-time (about 24 frames per second at 320x240 frame resolutions). The camera is fixed vertically on the wall of our experimental studio.

To obtain reasonable experimental result, we used several human models who have different appearance in face color, hair style and gender. All experiments were performed under the different illuminate condition (70~200 LUX).

As the result of experiments, the algorithm estimates articulated body position with small error fewer than 2.1cm in body center, 3.4cm in head center, 0.9cm in left and right hand respectably. All data could obtain in the real time (20frames/sec). Fig. 9 shows measured average error results.

IV. Conclusion

n this paper, we suggested a new approach of estimating human pose by analyzing morphological feature from the disparity map. From this approach, we can obtain the full body shape considering self-occlusion. It is simple and fast in comparison with other methods which satisfy real-time performance and accuracy of object tracking at the same time. Especially, the HCI on ubiquitous system will be appropriate application for this method.

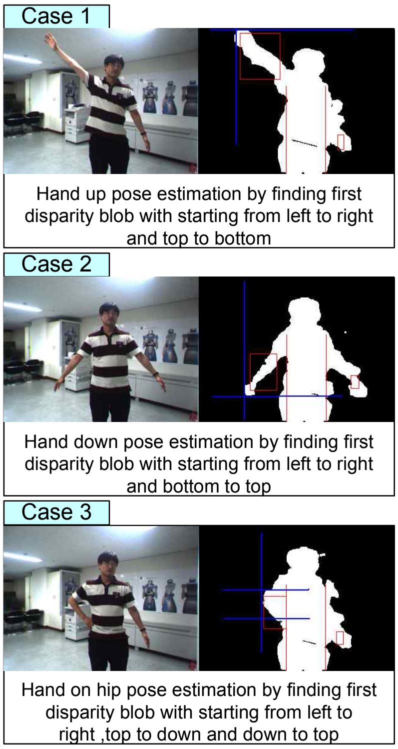

Figure 7. Morphological left hand detection case procedure in case of non self-occluded pose

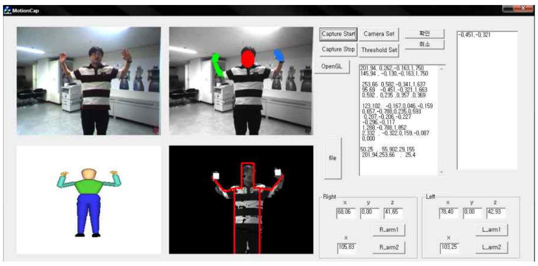

Figure 8. Experimental system setup and implemented human motion capture system.

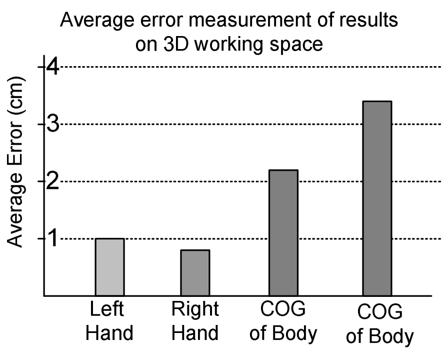

Figure 9. Experimental results; Average measurement of distance errors on each articulated object.

References

[1] E. Trucco and A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice Hall, New Jersey, 1998.

[2] T. Fong, I. Nourbakhsh and K. Dautenhahn, “A survey of socially interactive robots, Robotics and Autonomous Systems,” 42:143-166, 2003.

[3] M.K. Leung, Y.H. Yang, “First Sight: A Human Body Ouline Labeling System.” IEEE Trans. PAMI., 17(4):359-377, 1995.

[4] I. Kakadiaris, D. Metaxas, “Model-based estimation of 3D human motion,” IEEE PAMI, pp. 1453-1459, 2000.

[5] A. F. Lerasle, G. River, et al, ”Human Body Limb Tracking by Monocular Vision,” Scandinavian Conference, Volume 2,20-27 Sept. 1999 Page(s):716-721 vol.2.

[6] J. Yang, R. Stiefelhagen, U. Meier, and A. Waibel, “Real-time face and facial feature tracking and applications,” in Proc. of AVSP’98, pp 79-84, Terrigal, Australia, 1998.

[7] K. Konolige and David Beymer, SRI Small Vision System: Calibration Addendum to the User’s Manual Software version 3.2g, Videre Design, November 2004.

[8] K. Konolige and David Beymer, SRI Small Vision System: User’s Manual Software version 3.2g, Videre Design, November 2004.

[9] R. Tsai, “A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses,” IEEE Journal of Robotics and automation, Vol. RA-3, No. 4, 1987.

[10] Sewoong Jun, “Stereo Camera based Human Motion Capture by using Expectation Maximization Algorithm,” Dept. of Electrial and Electronics Engineering. The Graduate School of Yonsei Univ. 2006.