Real-time Upper Body Pose Detection using Stereo Vision ASIC

Jae-chan Jeong [1,2], Ho-chul Shin [1] , Dae-hwan Hwang [1]

[1] U-Robot Research Division , Electornics and Telecommunication Research Institue, Daejeon, Korea
(Tel : +82-42-860-6116;E-mail:channij@etri.re.kr)
[2]Department of Computer Software & Engineering , Korea University of Science and Technology, Daejeon ,Korea

Abstract

In this paper, we present our stereo vision ASIC and a real-time upper body pose detection algorithm. The developed an ASIC which can generate QVGA(320×240) 30fps stereo map mounted on embedded system. We also developed a real-time upper body pose detection algorithm which is executed in the embedded system. The developed hardware and detection algorithm can extract various human upper body poses at 30 fps. The computational complexity of proposed algorithm is quite low and it required CPU occupation under 60% in 350MHz embedded processor.

Keywords stereo vision; body pose detection; human-computer interface

1. Introduction

Human pose detection is very useful function for innovative human-computer interfaces and most of the human gesture information is generated at upper body. There have been many studies on human upper body pose detection, and we can classify them roughly into contact types and non-contact types [1]. In case of contact type, magnetic sensors or reflective markers can show quite high accuracy and reliability, but they are inconvenient because users must wear sensors or markers during the detection. In case of non-contact type, the image analysis methods can lighten the burden on users. On the other hand, the user gesture detections using image processing can be classified into single camera and multiple camera system. For single camera body pose detection [2]-[5], it has benefits that general movie or camera devices can be used. But in that case, as a single camera cannot provide distance information, accurate poses are hard to detect. There are various studies to solve this problem using stereo camera [6]-[11]. A stereo camera system can increase gesture detection result, but it requires high price stereo camera and additional high cost computational environment.

On the other hand, there are demands for low price and reliable stereo vision system in various fields, but existing stereo vision systems [12]-[14] are quite expensive for industrial products. As an example, many researchers are using the BumbleBee® system from PointGrey [12], but it has disadvantages such as high price and additional high power computer for stereo processing. To solve these problems, we developed a stereo vision processing ASIC which can provide QVGA (320x240) 30fps stereo map. We also developed a real-time upper body pose detection algorithm for this hardware system.

2. Stereo Vision Processing ASIC

As the stereo vision processing is repetitive calculation, we developed a stereo vision ASIC named as FALCON using parallel trellis dynamic programming [15]. This 200 million gate ASIC can produce QVGA 30fps distance map (Fig. 1,2). We also developed three base lines (6, 8, 12 cm) stereo CMOS camera. To evaluate the accuracy of the stereo ASIC, the error between the actual distance and ASIC distance map was compared. It shows error below than 3cm within 180cm distance (Fig. 3).

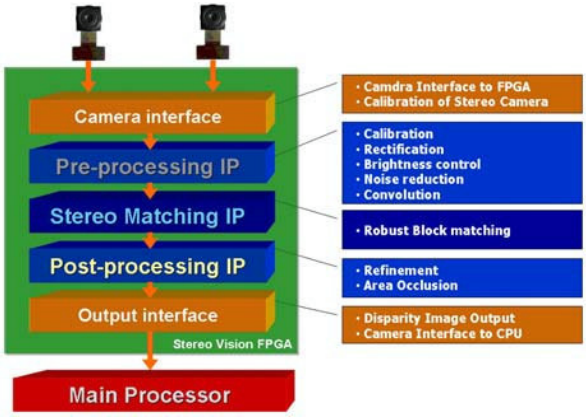

Fig. 1. Developed stereo vision ASIC structure

Fig. 2. Developed ASIC for stereo vision processing

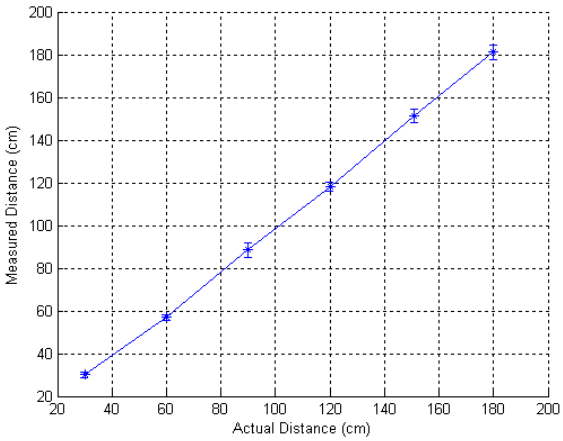

Fig. 3. Distance accuracy of the stereo vision ASIC

3. Real Time Upper Body Pose Detection

3.1. Face and Hand Detection

For real time upper body pose detection, we detected foreground skin regions at first. If they have reasonable sizes and positions, the regions are classified into face and hand. Based on this information, the shoulder and elbow positions are estimated (Fig. 4).

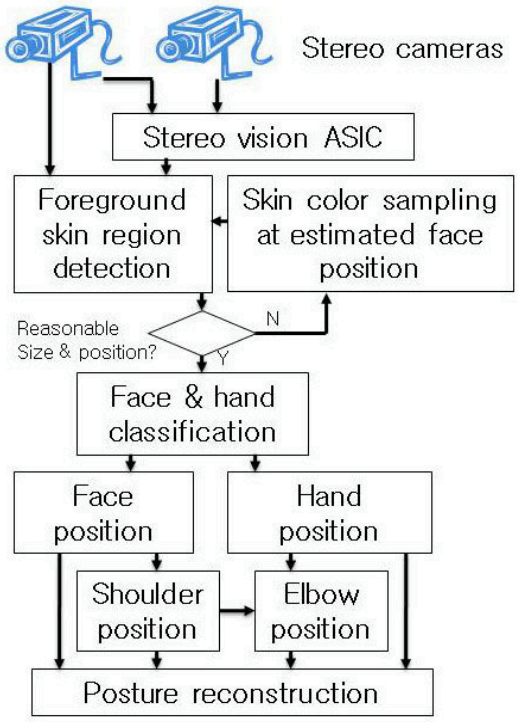

Fig. 4. Posture detection procedure

A skin color detection is widely used for face detection and gesture detection[16]. In this study, we used the chrominace component of the original images and adopted the elliptical boundary model [17] for skin region detection as shown in (1).

$$\Phi(c) = [c - \Psi]^T \Lambda^{-1} [c - \Psi] \qquad (1)$$
$$\Psi = \frac{1}{n}\sum_{i=1}^{n} c_i, \Lambda = \frac{1}{N}\sum_{i=1}^{n} f_i (c_i - \mu)(c_i - \mu)^T$$

Where c is the chrominance component vector, N is the total number of samples, f_i is the number of samples with chrominance c_i and μ is the mean of the chrominance vectors in the sampled data set. c_i is the sampled chrominance vector. Λ is the covariance matrix. The pixel chrominance c is classified as a skin pixel, if $\Phi(c) < \phi$, where ϕ is a threshold value chosen empirically as a trade-off between the true and false positives. Using this model and distance map, foreground skin regions were detected(Fig. 5).

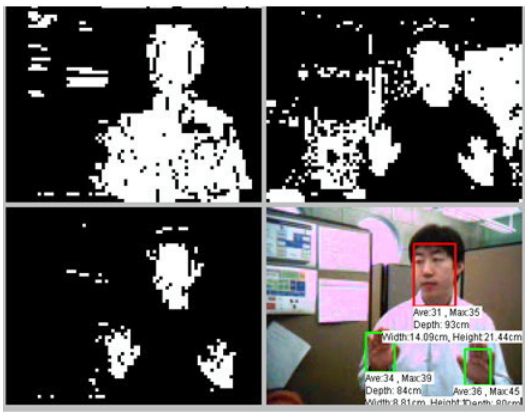

Fig. 5. Foreground skin region detection

Generally human faces have more horizontal patterns than hand, horizontal egdes can be used for face and hand classification. We calculated horizontal edges on detected skin reigeon with sobel mask. We defined the face scores which are multiplied skin region actual height by horizontal pattern count. The highest face score region was decided as face, and other regions are decided as hands. The propoesed algorithm can be applied to long sleeves. We are developing advanced algorithms for short sleeves or skin color clothes.

3.2. Shoulder and Elbow Position Estimation

We developed algorithm for shoulder and elbow position as follows. If we assume that users have standard Korean body size [18], The shoulder positions can be estimated from the face position and upper and lower arm lengths are estimated also. Once the shoulder and hand positions are detected, we have elbow candidates . As shown in Fig.6

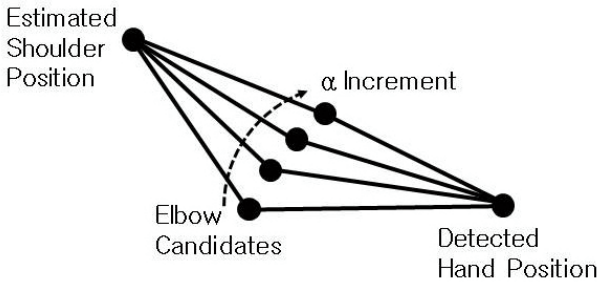

Fig. 6. Elbow Candidateds

As the α increases, the elbow candidates are changed as shown in Fig. 6. By minimizing the distance error between elbow candidate distances and stereo vision result, defined as $\left|p_e^z - D(p_e^x, p_e^y)\right|$, we can optimize the elbow position.

$D(p_e^x, p_e^y)$ is stereo map value of point p_e^x, p_e^y. After all, the face, shoulder, elbow, hand positions can be determined and upper body pose is reconstructed as shown in Fig. 7.

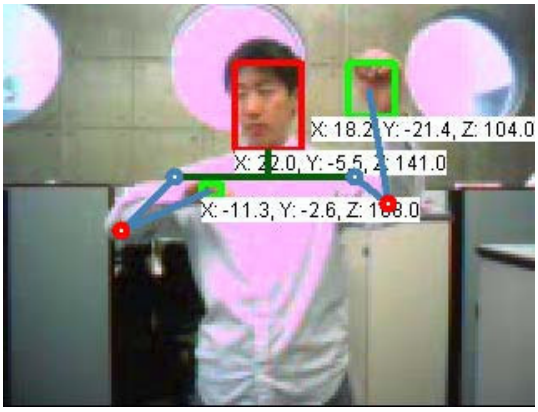

Fig. 7. Reconstructed upper body pose

4. Performances

In Fig. 8, various upper body pose detection results are shown. From the first column, it shows the original image, stereo result, side view, front view.

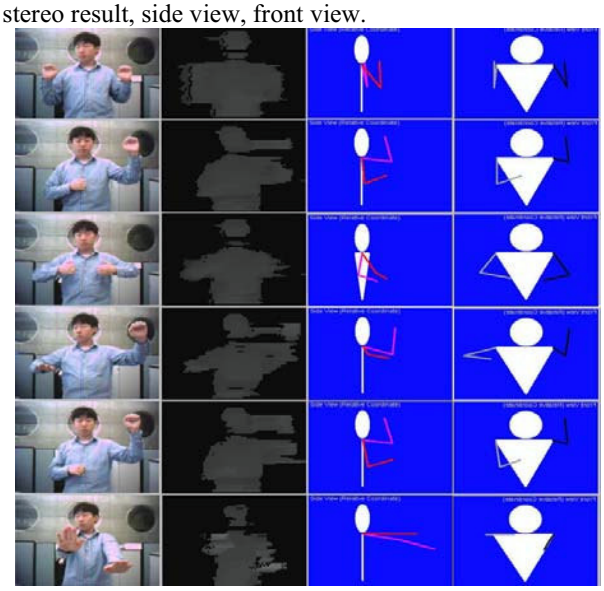

Fig. 8. Detected various poses

To evaluate the position accuracy, we compared the actual hand and elbow position using 3D position measurement device which has sub-millimeter accuracy, and detected hand and elbow position from developed system (Fig. 9). Three adult subjects were measured and their statures were

ranged from 167cm to 178cm. Because the elbow position is derived from hand position, it has more errors than hand. It shows errors about several centimeters. The detection rate was defined as the ratio of successfully detected frame and total frame. The developed upper body pose detection algorithm required under 60% CPU occupation for QVGA size, 30fps at 350MHz embedded system. There are few studies referring computational costs about upper body pose detection, but we can find out that our algorithm requires very low computational costs considering the result of [6], [8].

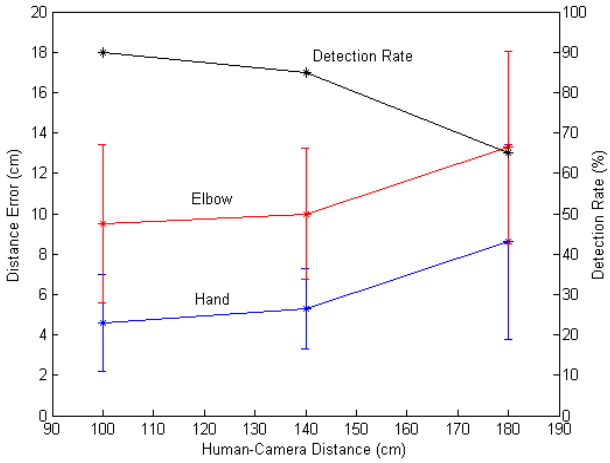

Fig. 9. Detection rate and position accuracy

5. Conclusion

In this study, we developed and verified a stereo vision ASIC and a real-time upper body pose detection algorithm. The developed system can detect various human upper body poses for 30 fps, shows about several centimeters error for hand and elbow position within 1.8m human-camera distance. Because the standard body size was assumed, the errors may be increased for small stature user or children. The developed algorithm requires very low computational cost below than 60% CPU occupation at 350MHz embedded processor. Because it requires low cost hardware, we hope our system can be widely applied to innovative human-computer interface, such as video games and remote controllers of various home appliances.,

References

[1] K. Takahashi, T. Sakaguchi, J. Ohya, “Remarks on a Real-Time, Noncontact, Nonwear, 3D Human Body Posture Estimation Method”, Systems and computers in Japan, v.31 no.14, 2000, pp.1-10

[2] R. Bowden, T.A. Mitchell , M. Sarhadi., “Non-linear statistical models for 3D reconstruction of human pose and motion from monocular image sequences”, Image and vision computing, v.18 no.9, 2000, pp.729-737

[3] M. Yamamoto, “A Simple and Robust Approach to Drift Reduction in Motion Estimation of Human Body”, Electronics and communications in Japan. Part 3, Fundamental electronic science, v.89 no.8, 2006, pp.39-52

[4] C.R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, “Pfinder: Real-Time Tracking of the Human Body”, IEEE transactions on pattern analysis and machine intelligence, v.19 no.7, 1997, pp.780-785

[5] M.W. Lee, I. Cohen, “Human Upper Body Pose Estimation in Static Images”, Lecture notes in computer science, v.3022, 2004, pp.126-138

[6] K. Nickel, R. Stiefelhagen , “Visual recognition of pointing gestures for human-robot interaction”, Image and vision computing, v.25 no.12, 2007, pp.1875-1884

[7] H.D. Yang, S.W. Lee, “Reconstruction of 3D human body pose from image sequences based on top-down learning”, Pattern recognition, v.40 no.11, 2007, pp.3120-3131

[8] K. Ogaki, Y. Iwai, M. Yachida, “Posture Estimation Based on Motion and Structure Models”, Systems and computers in Japan, v.32 no.4, 2001, pp.48-58

[9] M. Shimizu, T. Yoshizuka, H. Miyamoto, “A gesture recognition system using stereo vision and arm model fitting”, International congress series, v.1301, 2007, pp.89-92

[10] J. Mulligan, “Upper Body Pose Estimation from Stereo and Hand-face Tracking”, Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on 2005, pp.413-420

[11] H.D. Yang, S.W. Lee, “Reconstructing 3D Human Body Pose from Stereo Image Sequences Using Hierarchical Human Body Model Learning”, Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, 2006 v.3, 2006, pp.1004-1007

[12] http://www.ptgrey.com/products/bumblebee2

[13] J.I. Woodfill, G. Gordon, D. Jurasek, T. Brown, R Buck., “The Tyzx DeepSea G2 Vision System, ATaskable, Embedded Stereo Camera”, Computer Vision and Pattern Recognition Workshop, 2006 Conference on, 2006, 2006, pp.126-126

[14] http://www.videredesign.com

[15] H. Jeong, S. Park, “Generalized Trellis Stereo Matching with Systolic Array”, Lecture notes in computer science, v.3358, 2004, pp.263-267

[16] P. Kakumanu, S. Makrogiannis, N. Bourbakis, “A survey of skin-color modeling and detection methods”, Pattern recognition, v.40 no.3, 2007, pp.1106-1122

[17] J.Y. Lee, S.I. Yoo, “An elliptical boundary model for skin color detection”, Imaging science, systems, and technology; CISST'02, 2002, 2002, pp.579-586

[18] http://sizekorea.kats.go.kr