Depth Calculation by using Face Detection ASIC

Seung Min Choi, *Jae-chan Jeong , Jaeil Cho

Electronics and Telecommunications Research Institute
*University of Science and Technology

{ccsmm, jicho }@etri.re.kr
*channij@ust.ac.kr

Abstract - Real-time processing at mobile devices is always a hot issue. Although face detection and stereo matching, specially, are core elements for HRI application, they are one of the processes with heaviest loads at real-time processing. In this paper, we present novel methods of calculating disparity and detecting face with one chip, named Altair, which can detect 32 frontal faces/frame at 30 fps. Stereo image can be inputted into Altair, because it is designed to detect face on frame by frame and has two input ports. If the left image is inputted on N-th frame time and then the right image is inputted on (N+1)-th frame time, we can get face information of left and right images with one frame delay. Then we can calculate the disparity of face with this information and we can easily estimate distance from robot to human with it. When one or more faces are detected, we classify various cases to identify correspondent faces. We can detect faces and calculate the distances of them at 15 fps and can adjust the frame rate.

Keywords: HRI, HCI, face detection, stereo vision, disparity, real-time

I. INTRODUCTION

We introduce a novel method of calculating disparity of stereo camera with face detection chip. Because, real-time processing at mobile devices has been always a hot issue, many researchers have studied about speedy and light image processing algorithms for years. Although, face detection and stereo matching, specially, are core elements for HRI application, they are one of the applications with heaviest loads at processing.

In this paper, we present novel methods of calculating disparity and detecting face with one chip, named Altair, which can detect 32 frontal faces/frame at 30 fps. Stereo image can be inputted into Altair, because it is designed to detect face on frame by frame and has two input ports. If left image is inputted on N-th frame time and then the right image is inputted on (N+1)-th frame time, we can get face information of left and right images with one frame delay. Then we can calculate the disparity of face with this information and we can easily estimate distance from robot to human with it. When one or more faces are detected, we classify various cases to identify correspondent faces.

II. Stereo Camera and face detector

Disparity can be calculated by finding correspondence of stereo cameras. Fig.1 shows geometry of stereo camera system which is consists of two camera and field of object models. Because in fig.1 there are 4 triangles which have properties of triangle similarity, depth of disparity equations can be derived from them like table 1.

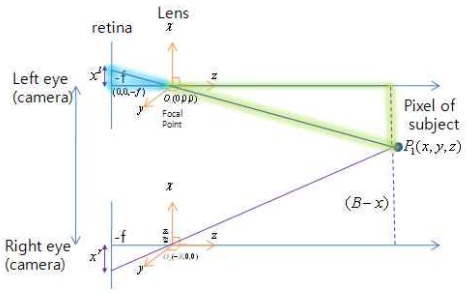


Fig.1. Geometry of Stereo Camera Model

Table 1. Equation of Stereo Camera Model

By triangle similarity of Left Camera System,

$$x^l : x = f : z \rightarrow xf = zx^l \rightarrow x^l = \frac{xf}{z}$$

By triangle similarity of Right Camera System,

$$x^r : B - x = f : z \rightarrow (B-x)f = zx^r$$

$$\rightarrow x^r = \frac{Bf - xf}{z} \rightarrow x^r = \frac{Bf}{z} + \frac{-xf}{z} = \frac{Bf}{z} - x^l$$

$$\therefore d = \frac{Bf}{z} \text{ ,or } z = \frac{Bf}{d}$$

$$\text{, where } d = x^l + x^r \text{ (absolute value)}$$

In generally finding correct disparity of stereo images is key point of stereo correspondence problem. So, many stereo vision algorithms are suggested over the years. According to [1], however, there are few real-time algorithms (over 30 fps) as shown in taxonomy.

Because of difficulty of real-time processing and huge power consumption for stereo matching, in some application where disparity of whole image pixel are not need, only disparity or depth of ROI(Region Of Interest) is important value in the camera image. For example, when a service robot has to keep a distance and serves some convenience to human, it only needs to know the distance from human to it. If it has a face detection function with its platform, it's easy to know the distance from human.

We present the way that calculates the distance with face detection chip. Fig.2 shows the outline of the novel method of calculating disparity of stereo camera with face detection chip. Two of camera image enters face detection unit in time division manner. Then face detector finds face and calculate the difference of center of face in left and right images.

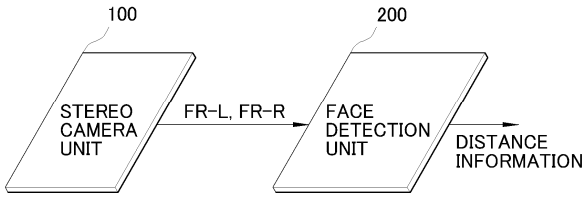


Fig.2. Outline of novel method.

We use the Altair as a face detection unit in fig. 3. Altair is the ASIC chip($1.2\times1.2cm^2$) for face detection that can detect 32 person per frame at 30 fps of QVGA(320x240) image[3]. It covers frotal and $\pm15^\circ$ slant face. The chosen method of face detection is the AdaBoost algorithm, which consist of MCT (Modified Census Transform) confidence

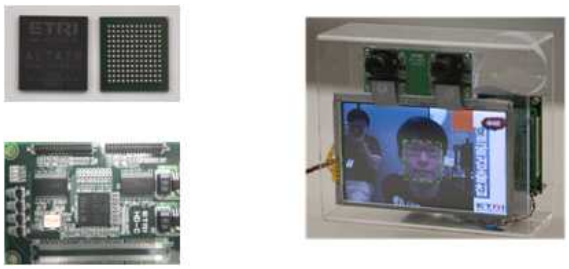


Fig.3. Altair (Frontal face detect ASIC)

III. Disparity Calculation with Face Detector

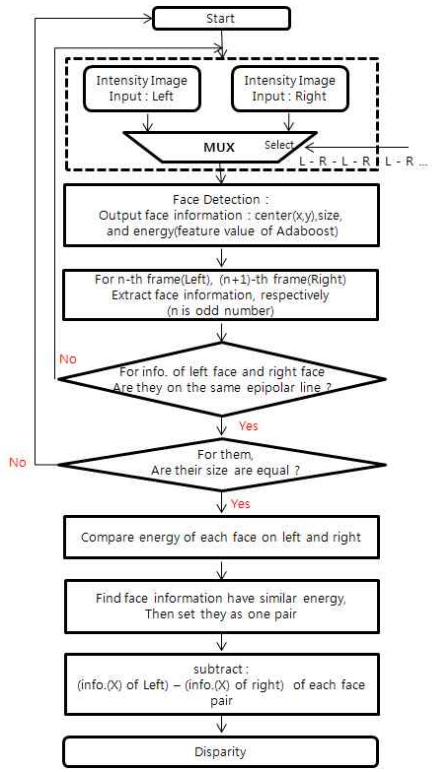
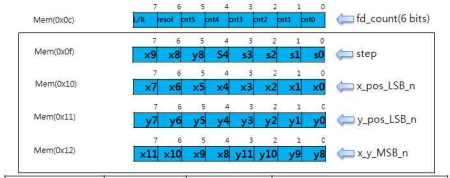


Fig.4. Flow Chart of our new method.

Fig.4 shows the flow chart of our novel method for calculating disparity of stereo camera with face detection chip. After start left and right camera images input into the multiplexer (MUX), then left or right image outputs from MUX in regular sequence by 'Select' signal. Then this interlaced image with frame order enters face detection block whose output is information of faces that consists of following table 2.

Table 2. Face information register of Altair



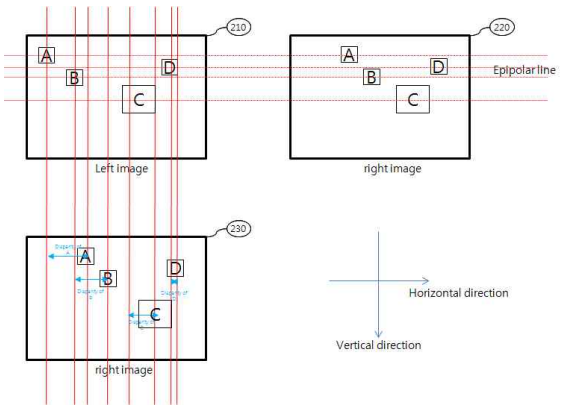
Register	range	Bit width	Bit
FD_count	0~31	5+1(L/R)	Fd_count(5:0)
Bits/frame		6	

Register	range	Bit width	Bit
STEP_n	0~20	5	step(4:0)
X_POS_n	0~639	10	step(7:6) & x_pos_LSB(7:0) or, x_y_MSB(7:4) & x_pos_LSB(7:0)
Y_POS_n	0~479	9	step(5) & y_pos_LSB(7:0) or x_y_MSB(3:0) & x_pos_LSB(7:0)
Bits/person		24	

After getting face information from Altair, next step is to compare information of faces on left frame and on right frame as shown in table 3.

Table 3. Comparing face and Calculating disparity

item	Information	note
On the same epipolar line ?	'x' coordinate of center point	Vertical direction
Has the same size of face	Size of face	
Has the similar energy value?	Sum of Confidence	Sum of feature of Adaboost algorithm (1/probability)
Calculate Disparity	'x' coordinate of center point	Horizontal direction



VI. Experiment

We capture two sets of stereo image with system in fig. 4. One is the continuous image set where a human moves around camera as shown in fig. 5

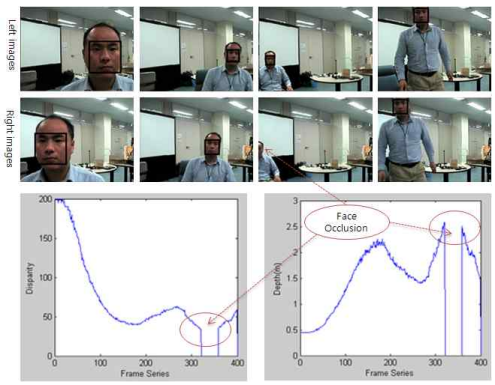


Fig.5. depth calculation for moving human

A human is near camera at first and move far from it then comes to near again. Because face was not detected for a while at around about 350-th frame, face occlusion occurred and disparity and depth could not be estimated. Depth can be converted from disparity by using following equation which is modified version of table 1.

We multiply the constant 'k' to depth, because B(baseline),f(focal length) and wpp(width per pixel) may have measurement error, and it's hard to know correct parameters for the combination of lens and CMOS sensor used in this experiment('k' is a experience value).

$$Depth = \frac{B \times f}{d} = k \times \frac{B \times f}{disparity \times wpp}$$

$where, our\ system\ has$

$B = 0.24m$

$f = 0.0045m$

$wpp = 0.0000123m$

$k = 0.85$

Another test set consists of 7 frames where a human has a paper written the real measured distance from camera to him as shown in fig.6. The distances are 0.5m ~ 3.5m with step of 0.5m. After gaining disparity with this second test set by our new method, we compared it to real measured distance as shown fig. 7. and table 4. We can detect faces and calculate the distances for human in range of 0.5m ~ 3.5m with 0.5~ 8.9% error

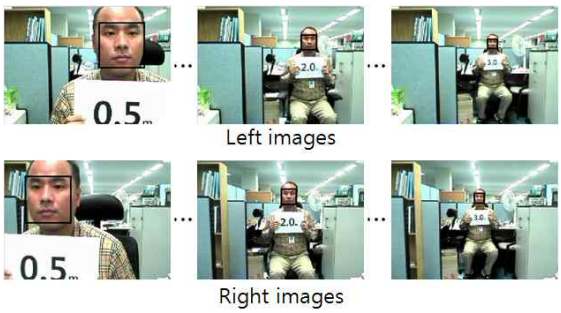


Fig.6. test image set with real distance value (0.5~3.5m)

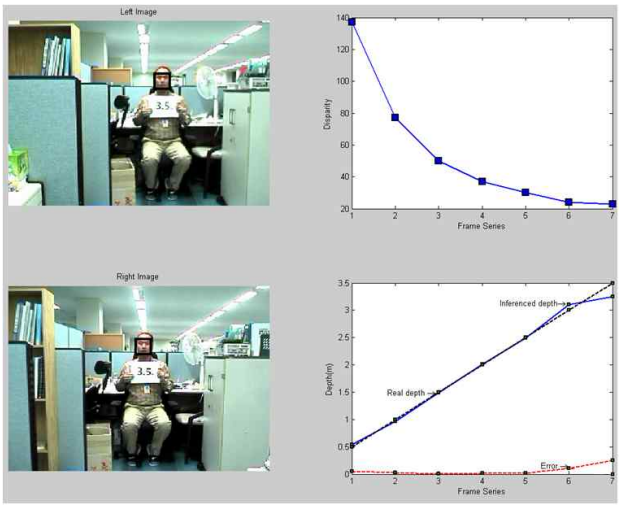


Fig.7. our method's accuracy of depth calculation

Table 4. Result of Depth calculation
(average of tests in 5 times)

Frame	Depth(m) (Inference)	Depth(m) (Real Measurement)	Error(m) (Difference of two)
1	0.5448	0.5	0.0448(8.9%)
2	0.9693	1.0	0.0307(3.1%)
3	0.1497	1.5	0.0073(0.5%)
4	2.0171	2.0	0.0171(0.9%)
5	2.4878	2.5	0.0122(0.5%)
6	3.1098	3.0	0.1098(3.7%)
7	3.2450	3.5	0.2550(7.0%)

Because Altair has two ports for input image and stereo camera inputs are multiplexed with time division manner, frame rate for depth result is half (15 fps) of Altair's that. By configuring Altair chip, we can adjust the frame rate of face detection and depth calculation.

V. CONCLUSION

We can calculate the disparity of face with this information and we can easily estimate distance from robot to human with it. When one or more faces are detected, we classify various cases to identify correspondent faces. We can detect faces and calculate the distances for human in range of 0.5m ~ 3.5m with 0.5~ 8.9% error

We are planning to do experiment for the case where there are many people a frame

Acknowledgement

This work was supported by the R&D program of the Korea Ministry of Knowledge and Economy (MKE) and the Korea Evaluation Institute of Industrial Technology (KEIT). [Project KI001813, Development of HRI Solutions and Core Chipsets for u-Robot].

REFERENCES

[1] Scharstein, Daniel and Richard Szeliski., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms", International Journal of Computer Vision. 2002

[2] Jiho Chang, Seung Min Choi, at el, " Design and Implementation of Real-Time stereo vision ASIC and embedded system", IEEK 2010 Summer Conference, CFP-326, June, 2010.

[3] Seung Min Choi, at el, " Implementation of Robust Real-time Slant Face Detection ASIC", IEEK 2010 Summer Conference, CFP-326, June, 2010.