The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Automatically Available Photographer Robot for controlling Composition and taking pictures

Myung-Jin Kim, Tae-Hoon Song, Seung-Hun Jin, Soon-Mook Jung,
Gi-Hoon Go, Key-Ho Kwon and Jae-Wook Jeon

Abstract— **Recent advances made in IT technology has given much impetus to the development of multimedia devices. The digital camera is such a multimedia device. It has made much progress and become very popular, as most people now own a digital camera or cell phone with camera features. People often take photographs in everyday life. Professional photographers often take photographs of travel destinations, banquet halls or parties.**

In this paper, we propose an autonomous robot photographer capable of taking pictures and thus replacing photographers. This photographer robot can detect direction based on the human voice. It can control composition based on skin color detection to snap the picture.

I. Introduction

People have recently become interested in photography due to continuous advancements in IT technology and widespread use of digital cameras. Nowadays, most people have digital cameras or cell phones with cameras. This makes cameras more popular. Furthermore, more people are interested in improving the quality of their photos. People have been actively exchanging information and posting pictures on internet community sites.

There has been recent research on improving photography skills. A basic principle for taking pictures is the rule of thirds [1]-[3] that equally divides each height and width into three parts. The subject being photographed is located on one of the intersection points. The rule of thirds can bring a sense of stability and a sense of beauty to photography. Another method is the one tenth outer boundary line principle that divides each height and width into ten equal parts; the subject being photographed is not located in the tenth part for picture stability. In this paper, the photographer robot uses the rule of thirds.

Research on photography includes detecting dissection lines in environmental portrait photos [4]. In this paper, we propose a method for detecting composition lines and controlling robot position. First, we detect the composition line in photos. Second, we check if the lines detected cross human objects. If so, these detected lines are defined as unnecessary composition lines in an environmental portrait photo. Finally, we move the robot by the position of the unnecessary composition line.

Research on automatic photographer robot has advanced. A photographer robot is now capable of taking pictures during travel, and in banquet halls or parties, acting as a professional photographer. For example, SONY's 'party shot' [5] can operate as a simple photographer robot. This product can be set on a table, rotating in place until it detects a human face. If it finds a human face, then the camera takes a snap shot and then resumes face detection. Fig. 1 shows Party-Shot. In this paper, we propose another photographer robot that can automatically change the robot location.

A number of studies on photographer robots exist. For example, a typical study in this field is William. D. Smart's photographer robot [6]-[8]. This paper proposes a photographer robot that consists of two cameras, robot vehicle and PC. The first camera is used to capture an image for vision processing, while the second camera is used to take the final picture. The PC performs vision processing using the image captured by the first camera. The vision processing result is used to control the robot vehicle. When, this process is completed, the second camera takes a picture. Finally, the PC uploads the final picture to the Internet.

Research on methods of taking pictures and photographer robots has recently made much progress [9] [10]. In this paper, we propose a robot photographer capable of detecting direction via human voice recognition, extracting human objects using skin color detection and extracting composition

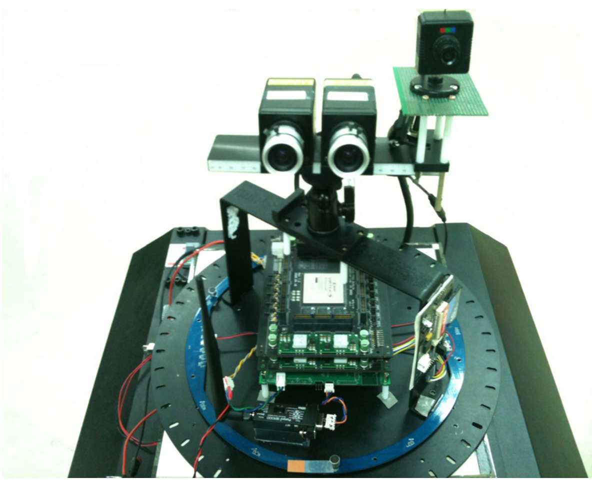
Fig. 1. 'Party-Shot', SONY, Japan

Manuscript received March 10, 2010. This research was supported by MKE, Korea under ITRC NIPA -2010-(C1090-1021-0008).

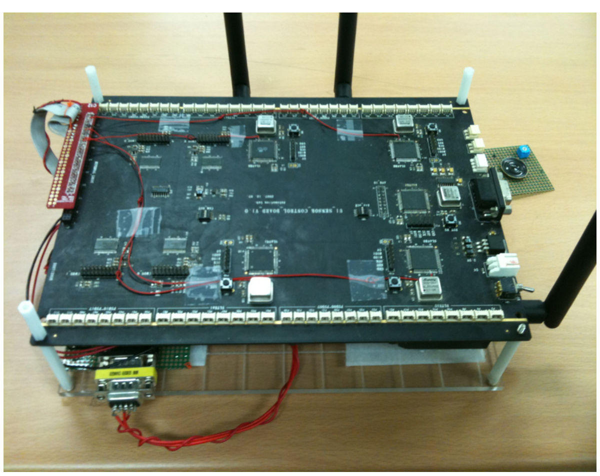
Myung Jin Kim, Tae Houn Song, Seung Hun Jin, Soon Mook Jeong, Gi Hoon Go and Key Ho Kwon are with the School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea (e-mail: {darkkmj, thsong, coredev, kuni80}@ece.skku.ac.kr, tokogi@hotmail.com, khkwon@yurim.skku.ac.kr).

Jae Wook Jeon is with the School of Information and Communication Engineering, Sungkyunkwan University, Suwon, Korea (e-mail: jwjeon@yurim.skku.ac.kr; phone: +82-31-290-7129; fax: +82-31-299-4921).

(a)



(b)



(c)

Fig. 2. Figure of each part of robot. (a) Vehicle Robot(NT-Giant-II) and (b) Wireless CCD Camera and Direction Detection System and (c) Main Controller Board.

lines across the human object. The PC then controls the photographer robot and takes a final photo using human object information and composition line information across the human object. To summarize, our photographer robot takes a picture using the basic principles of taking pictures, skin color detection and composition line detection.

II. System Configuration

A. Vehicle Robot

Our photographer robot uses the NT-Giant-II, a four-wheel-drive vehicle, as a vehicle robot. The vehicle robot is comprised of four DC motors and two motor drivers. Fig. 2 (a) shows the NT-Giant-II used with our photographer robot. The DC motor specs are as follows. Input voltage is 24VDC, power consumption is 300W, and motor torque is 9.74(kg-cm). Four DC motors and a battery are installed on the bottom of the NT-Giant-II robot. Two motor drivers and one robot control board are installed on top of the NT-Giant-II robot. The robot control board has firmware to control the NT-Giant-II. It communicates with the main control board using a wireless Bluetooth serial module. Our photographer robot can easily execute actions forward, backward, rotating left and rotating right. It can also control speed.

B. Direction Detection FPGA

A module to detect direction by human voice recognition uses the Real-time Sound Source Localization System based on FPGA [11]. This FPGA system was initially performed by Jin Seung-hoon, Kim Dong-kyun and other researchers in our laboratory.

Three analog microphones are located in a circle at intervals of 120 degrees. If the microphone recognizes human voice, then analog data are converted to digital sampling data using the mini sampling board. Then, digital data are sent to the FPGA board. In FPGA, digital data received at the mini sampling board is used to detect direction. The result of direction detection is sent to the main control board using a wireless Bluetooth serial module that is connected to the serial port of FPGA and uses serial communication. A real-time Sound Source Localization System recognizes human voice and detects direction. The result data are then changed into a degree form. That is, 0~+180 or 0~-180. The FPGA generates result data using a pre-defined protocol and sends to the main control board. For example, <+180> or <-010>. These processes are performed in real time.

C. Wireless CCD Camera

We use a wireless CCD camera in our photographer robot. It consists of a CCD camera and receiver. Fig. 2 (b) shows our wireless CCD camera. CCD camera specs are as follows. It has 2.4GHz four channels with 240,000 pixels available. The power consumption is 250mA, input voltage is 5VDC and the

RX-2400 is used as a receiver. Wireless communication is performed between the CCD camera and RX-2400. RX-2400 is connected to the master PC. The PC cannot directly use the output data, as the output from the receiver is a video signal. Therefore, we use a video signal-to-USB signal converter between the RX-2400 and master PC. Basically, we use one camera for vision processing, robot control and taking pictures. Therefore, for high resolution images, we will use two cameras for vision processing and snapping the result picture.

D. AVR Main Control Board

We use ATmega128 at the main control board to control the robot and to communicate with the FPGA used to detect direction. Three ATmega128 processors are used to communicate between the FPGA, robot control board and master PC. Fig. 2 (c) shows our main control board consisting of five ATmega128 processors. However, in this paper, we use only one master processor and two slave processors.

The ATmega128 master processor is connected to the PC through the main control board. Two slave processors are connected with the robot control board and direction detection FPGA using a wireless serial module. The ATmega128 master processor sends instruction from the master PC to each slave processor using SPI communication [12]. Then, each slave processor communicates with the robot control board and FPGA using serial communication. If the instruction received from the master processor is completed, the slave processor sends the result data to the master processor using SPI communication. If the master processor receives the result data from the slave processor, then the master processor sends the result data to the master PC. These processes represent the communication flow between the main control board, master PC, robot control board and FPGA. Fig. 3 shows the system configuration and information about communication between each module.

E. Master PC

The specification of the master PC used in our experiment is detailed as follows.

- CPU : Athlon 64 X2 Dual Core 3800+ 2.01 GHz
- RAM : 2.5GB
- OS : Windows 7 Ultimate K

The master PC receives the result data from the direction detection FPGA on our photographer robot. Next, the master PC sends instruction to the main control board to control NT-Giant-II using serial communication. The master PC executes image processing using the image data from the wireless CCD camera on our photographer robot. Image processing can be divided into two parts; skin color detection for human object extraction and line detection for composition line extraction. Next, the master PC predicts the route for our photographer robot using the information extracted. It transfers the instruction to control the photographer robot to the main controller using serial communication. When robotic control is complete, the master PC snaps the resulting picture using the wireless CCD camera on the photographer robot.

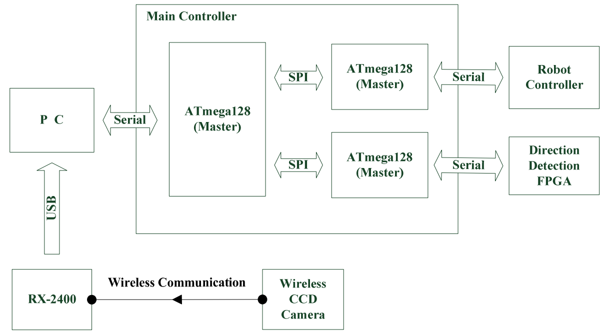


Fig. 3. System Configuration

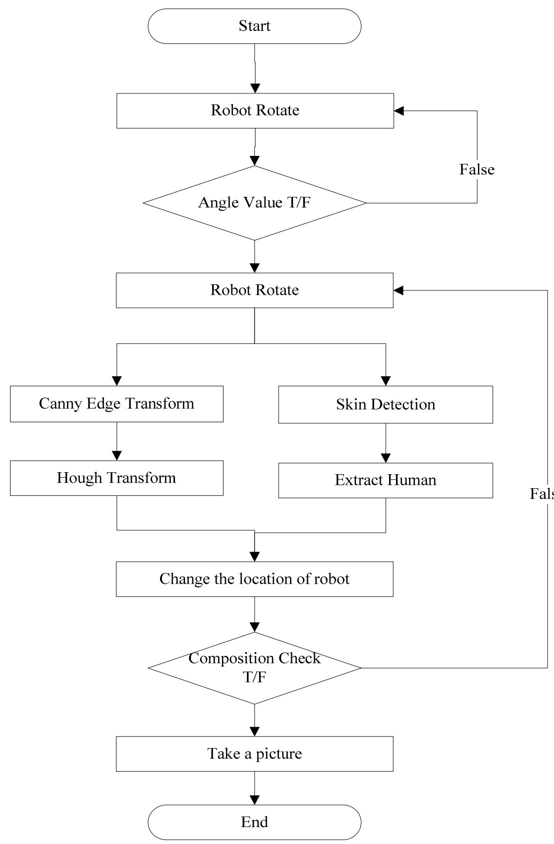


Fig. 4. System Operation Flow Chart

III. SYSTEM OPERATION FLOW

We have explained the function and operating principle of each module of our photographer robot. Next, we explain the overall system operation sequence. First, our photographer robot rotates to the direction detected by direction detecting FPGA using human voice. Then, the master PC performs vision processing to detect a human object and unnecessary

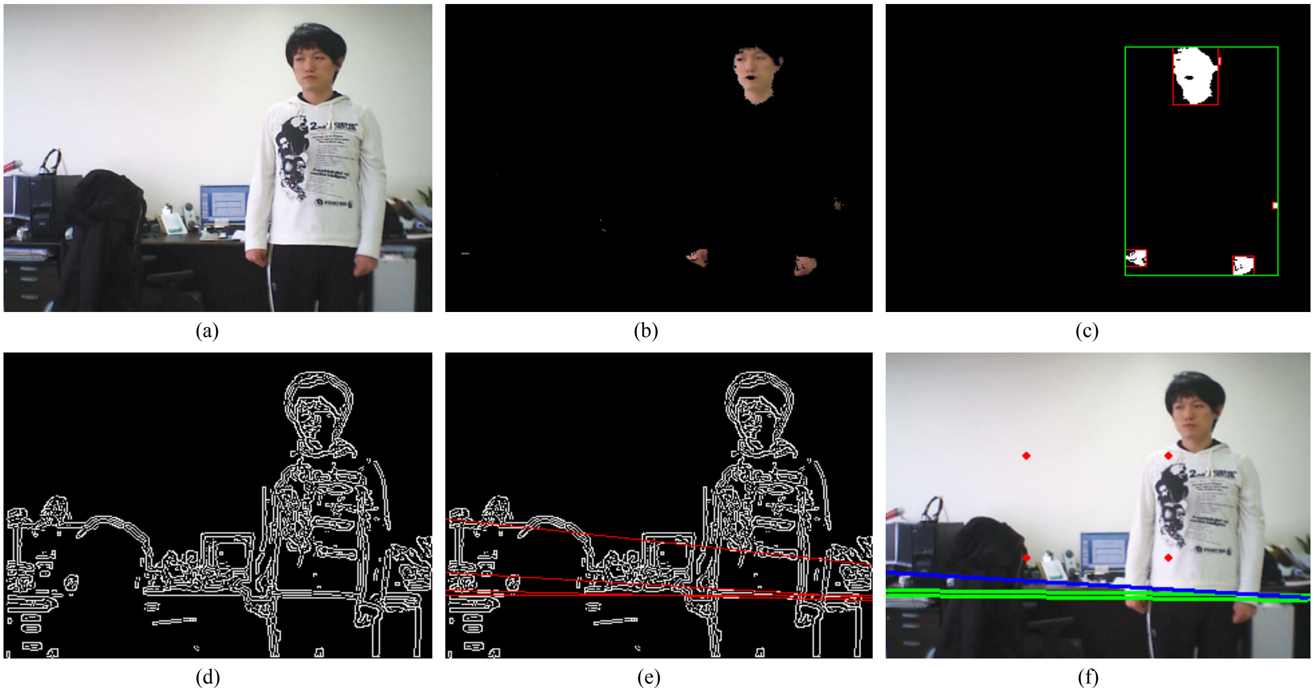


Fig. 5. Result image of each execution image processing. (a) Original image and (b) Skin detection result image and (c) Human detection result image and (d) Canny function execution result image and (e) Hough function execution result image and (f) Line detection result image.

composition lines using the image data from a wireless CCD camera. This resulting data are used to calculate the position of the photographer robot. For example, our photographer robot could locate the point at which the human object is located according to the rule of third. The resulting information is used to reject the line located across the neck or waist of the human. At this point, direct control of the robot is performed on the main control board. After deciding the position of our photographer robot, the master PC captures the image from the wireless CCD camera and repeats the entire process. Fig. 4 depicts the flow chart of the system operation.

IV. IMAGE PROCESSING

A. Human object extraction

We use the skin color detection method [13] to extract the human object from the received image. Studies on skin color detection provide many alternative methods and algorithms. The algorithm we need must be able to rapidly execute skin color detection and have high accuracy skin color detection of moving objects. Accordingly, we used the fusion Gaussian algorithm introduced in T. C. Pham's research [14]. As previously stated in this paper, this method is capable of detecting color by comparing data to the skin color learning table by the pixel.

Human object detection is as follows. First, we have to prepare a learning table that includes a variety of skin color data using a Gaussian mixed modeling algorithm. Second, we compare the prepared skin color learning table to the original image by pixel. If the compared result is greater than the predefined threshold value, then it is defined as skin color and extracted. Fig. 5 (a) shows the original image, and Fig. 5 (b) shows the resulting image of skin color detection. This skin color detection result image shows that it has not only detected human skin, but also a similar background color. Due to this noise, we calculate the size of each connected component, and we then reject the component regarded as noise due to its small size.

The labeling step is next. We performed component detection that rejected some noise to labeling using the Blob function of OpenCV. Each connected component had another label. If labeling is completed, then the labeled component is analyzed. We choose one connected component that is located in a higher position and has higher density and larger size than other similar components. We define it as a human head. We consider the position of the neck as the bottom of the human head. We also decide the location of the waist using a scale table. The scale table includes the value of the waist's height by each of several human face sizes.

We define the human object that consisted of the human face and the value of the calculated waist location. Fig. 5 (c) shows a result image of detecting a human object. We can change the location of our photographer robot by the rule of third using the human object.

B. Composition line extraction

This stage detects the composition line and determines the position of the photographer robot using line information and human object information [15]. For this stage, we use line detection functions from OpenCV 2.0 (Open source

Computer Vision library) that easily provides a variety of functions for vision processing. The entire process is explained below.

First, we transform the original image into a gray image. Second, we perform the canny conversion function at OpenCV for edge detection. The canny conversion function detects the point that sharply changes brightness in the image. There is a variety of the canny conversion methods, but we use the canny conversion method using prewitt mask. This method has several advantages. It performs conversion at a relatively fast speed and detects fewer edges than other methods, because it gives less weight to brightness changes. Since we need is 1~4 efficient lines as the composition line, we could detect more results using the canny conversion method. Fig. 5 (d) shows the canny function execution result image of the original image.

Next, we perform the Hough conversion function included in OpenCV using the canny conversion execution result image. Fig. 5 (e) shows the Hough function execution result image. The Hough conversion function can detect lines that meet the conditions in the image. And we executed additional detected line filtering. Therefore, we can almost detect lines in the current input image.

The resultant image has unwanted lines in addition to the lines we need. Therefore, we must perform efficient line selection via the Hough conversion result image. The outputs of Hough conversion are the start and end points of the line. However, if the length of the detected line is longer than the range of the x, y pixel in the currently used image, then the resultant point (start point, end point) also has a value that is out of the range we used. We used an input image size of 640x480. If the resultant point's value is out of this range, then line detection will encounter difficulties. Therefore, we converted the points with values in this range to avoid difficulty.

$$y = \frac{(y_2 - y_1)}{(x_2 - x_1)} x + \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1} \qquad (1)$$

Converting the point process proceed as follows. First, we calculate the slope of the line and the value of y when x is 0 using a start point, end point and line equation (1). Second, we calculate the converted points. As the detected line can only have contact with the y axis or x axis or both x axis and y axis, we calculate x when y is 0 or 240, or we calculate y when x is 0 or 320, according to each case. Finally, we calculate the points within the range we want.

Next, we group several efficient lines that have similar slope and position. We repeat this process until we have 1~4 lines remaining. Finally, we can extract 1~4 efficient lines. Fig. 5 (f) shows the resultant image. Red points show the intersection points of the rule of thirds.

We examine these final detected 1~4 lines for a line across the human neck or waist. At this point, we use the position value of the neck and waist that is defined during the human object extraction step. For example, the position value of the neck is defined under the position of the face. If the hand

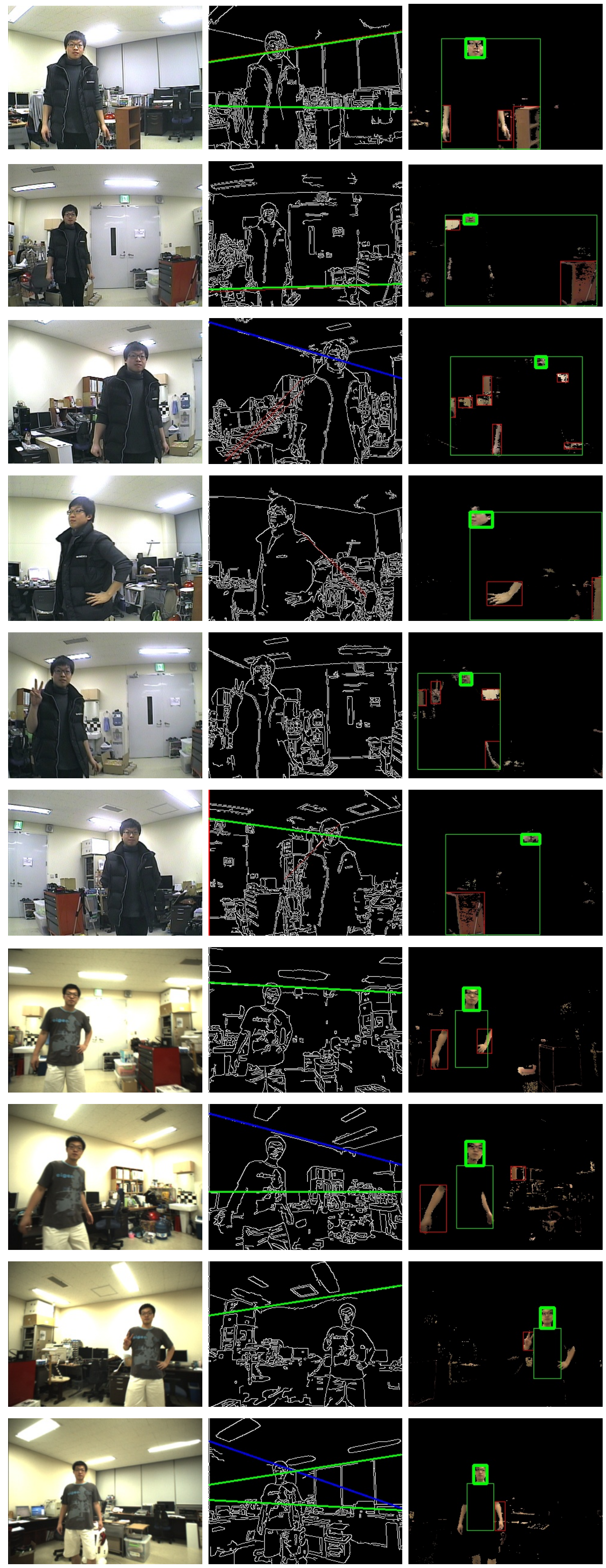


Fig. 6. Experiment result pictures. Left picture is result image and middle picture is composition line detection result image. Right picture is human detection result image.

position detected in the human object extraction step is located in a predefined range, then the position value of the waist is defined as a position of the hand. If the hand is not located in the range of predefined values of y, then we cannot define the waist position using the position of the hand. In this case, we define the waist position using predefined position value data defined according to the position of the human's head. In addition, if the object defined as a hand is not detected, then we defined the position of the waist using predefined position value data defined according to the head position.

If a line that crosses over the neck or waist exists, then the robot is moved to the other position to reject this line using the line slope and position value.

V. Experiment & Result

We tested our robot in the laboratory to evaluate the proposed photographer robot. Our photographer robot could correctly take many pictures. Our photographer robot took almost 100 pictures. We evaluated these resulting images using the rule of third and unnecessary composition line.

Eleven of these pictures were unwanted. These pictures include several bad points, in which the human is not located according to the rule of the third point, or they included a composition line that crosses over the neck or waist. Fig. 6 shows several resulting pictures. The middle pictures show the composition line detection results. In several situations, the composition line is not detected. However, we could find the composition lines we need in most instances.

Pictures in the right column show the human detection result. The bold rectangle defines the human face in the pictures. The thin rectangle is defined as human. We used the skin detection method; thus, it defined skin in situations in which the background has a color similar to skin. But in most instances, we could find the face exactly in the pictures. Therefore, we could use the position of the face and composition line information to control our photographer robot. Therefore, we mostly obtained pictures that satisfied our standard. In Fig. 6 the left column shows our final pictures. This shows they satisfied the rule of third and did not include composition lines that crossed over the neck or waist.

VI. Conclusion

The proposed photographer robot successfully snapped many pictures that satisfied our standards. This research discusses how the photographer robot successfully takes pictures using our scenario. However, our photographer robot must still improve upon its many weaknesses. We will improve the photographer robot, for example, by developing a driving method for our robot and improved methods of human object detection. We plan to develop picture capture standards and to enhance our photographer robot so it that can have greater interaction with humans.

Acknowledgment

"This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency" (NIPA-2010-(C1090-1021-0008))

References

[1] B. Gooth, E. Reinhard, C. Moulding, and P. Shirley, "Artistic composition for image creation", *Eurographics Workshop on Rendering,* 2001.

[2] B. Jenkins, B. L. Evans, "In-Camera automation of photographic composition rules", *IEEE Trans. Image Process,* vol. 16, no. 7, pp.1807-1820, Jul. 2007.

[3] Cavalcanti, R. M. C, Gomes, H, & Guerra, W. "Towards automating photographic composition of people", In *IASTED international conference on visualization, imaging, and image processing*, pp. 25-30, 2006.

[4] C. T. Shen, J. C. Liu, "Towards intelligent photo composition automatic detection of unintentional dissection lines in environmental portrait photos", *Expert Systems with Applications*, volume 36, Issue 5, July 2009, pp 9024-9030.

[5] SONY, "Party-Shot", *Internationale FunkausteLLung(IFA 2009),* Berlin, German, Sep 4-9, 2009.

[6] Z. Byers, M. Dixon, K. Goodier, C. M. Grimm, W. D. Smart, "An Autonomous Robot Photographer", *Paper presented at the international conference on Robots and Systems (IROS 2003)*, Las Vegas, Nev, October.

[7] Z. Byers, M. Dixon, C. M. Grimm, W. D. Smart, "Say Cheese! Experiences with a Robot Photographer", *AI Magazine* Volume 25, Number 3, 2004.

[8] W. D. Smart, C. M. Grimm, M. Dixon, Z. Byers, "(Not) Interacting with a Robot Photographer", *Proceedings of the AAAI Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*, March 2003

[9] J. Campbell, P. Pillai, "Leveraging limited autonomous mobility to frame attractive group photos", in *Proc. Of the 2005 IEEE Int. Conf. on Robotics and Automation (ICRA '05)*, Barcelona, Spain, 2005, pp. 3407-3412.

[10] H. S. Ahn, D. H. Kim, J. Y. Lee, S. Y. Chi, K. K. Kim, J. S. Kim, M. S. Hahn, and H. S. Kim, "A Robot Photographer with User Interactivity", *Proceeding of the IEE/RSJ International Conference on Intelligent Robots and Systems*, pp 5637-5643, Beijing, China, October 2006.

[11] S. H. Jin, D. K. Kim, H. S. Kim, C. H. Lee, J. S. Choi, and J. W. Jeon, "Real-time Sound Source Localization System based on FPGA", *IEEE International Conference on Industrial Information(IEEE INDIN)*, Daejeon, Korea, 2008.

[12] T. H. Song, J. H. Park, S. M. Jung, K. H. Kwon, and J. W. Jeon, "Embedded Robot Operating Systems for Human-Robot Interaction", *Asia-Pacific Conference on computer Human Interaction(APCHI)*, Incheon, Korea, 2008.

[13] V. Vezhnevets, V. Sazonov and A. Andreeva, "A Survey on Pixel-based Skin Color Detection Techniques", *Proc. Of Graphicon*, pp. 85-92, 2003.

[14] T. C. Pham, X. D. Pham, D. D. Nguyen, S. H. Jin and J. W. Jeon, "Dual Hand Extraction Using Skin Color and Stereo Information", *IEEE International Conference on Robotics and Biomimetics*, Bangkok, Thailand, Feb 21-26, 2009.

[15] J. Y. Hwang, D. S. Lim, D. W. Paik, "A Straight Line-Based Distance Measure to Compute Photographic Compositional Dissimilarity", *First International Conference, Future Generation Information Technology, FGIT 2009*, Jeju Island, Korea, December 10-12,2009