Vision System of Facial Robot SHFR- III for Human-robot Interaction

Xianxin Ke, Yujiao Zhu, Yang Yang, Jizhong Xing and Zhitong Luo
Department of mechanical and electrical engineering, Shanghai University, NO.149 Yanchang Raod, Shanghai, China

Keywords: Facial Expression Robot, Visual System, Human-robot Interaction.

Abstract: The improvement of human-robot interaction is an inevitable trend for the development of robots. Vision is an important way for a robot to get the information from outside. Binocular vision model is set up on the facial expression robot SHFR- III, this paper develops a visual system for human-robot interaction, including face detection, face location, gender recognition, facial expression recognition and reproduction. The experimental results show that the vision system can conduct accurate and stable interaction, and the robot can carry out human-robot interaction.

1 INTRODUCTION

The population aging is more and more serious in today's society, housekeeping and family nursing in the family are more and more demand for the robot. If robots can naturally communicate with people and express feelings at work, they will be better to help people in the psychological and physiological. Psychologists' studies show that only 7% of information is transferred by spoken language, while 38% is expressed by paralanguage and 55% is transferred by facial expressions. It can be seen that, facial expression robot continue to enter the people's lives, which will be an inevitable trend.

Based on the static or dynamic human face image, the robot can determine the relative position of them, which will be widely used in the future. The robot can control the robot's position and select the appropriate address and different communication methods, according to the information of the position, the expression and the gender. The visual research can provide personalized human-robot interaction mode for the robot, which is helpful to build complex and intelligent human-robot interaction system.

With the improvement of the humanoid robot human-robot interaction attention, more and more organizations begin to participate in them. Foreign studies carried out earlier, the early classic is infant Kismet robot of MIT(Massachusetts Institute of technology) that can be a natural and intuitive interaction, whose vision system are consist of face detection, motion detection and skin colour detection. KOBIAN-RII (Trovato et al., 2012), a relatively mature research by the Waseda University in Japan, can achieve dynamic emotional expression. And Italy's programmable humanoid robot iCub (Parmiggiani et al., 2012), mainly used for children's cognitive process.

Based on the facial robot SHFR- III, this paper develop visual system of robot, including face detection, face location, gender recognition, facial expression recognition and reproduction, which have great significance for the further development of human-robot interaction system.

2 FACIAL ROBOT SHFR- III

The platform is the facial robot SHFR- III. The robot's head is divided into four parts: the eyebrows mechanism, the eye mechanism, the jaw mechanism, the neck mechanism. It have 22 degrees of freedom, and each degree of freedom is controlled by a steering gear. PC sends a command to down-bit machine FPGA via serial communication. After getting instruction information the next crew sent PWN signal to the steering engine, and then the steering engine control rotation Robot expression. Face robot head and control box as shown in Fig. 1.

The binocular vision model of facial robot consists of two parts: hardware and software. The software is composed of the host computer vision processing algorithm, the target location algorithm

Ke, X., Zhu, Y., Yang, Y., Xin, J. and Luo, Z.
Vision System of Facial Robot SHFR- III for Human-robot Interaction.
DOI: 10.5220/0005994804720478
In *Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2016) - Volume 2*, pages 472-478
ISBN: 978-989-758-198-4

and the gender expression recognition algorithm. Hardware system includes image sensor, image acquisition card and PC, the system block diagram is shown in Fig. 2

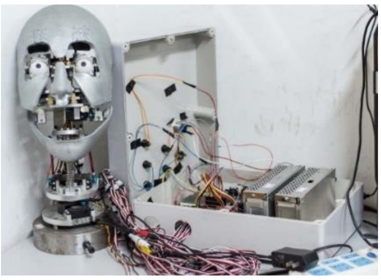


Figure 1: Facial Robot SHFR- III platform.

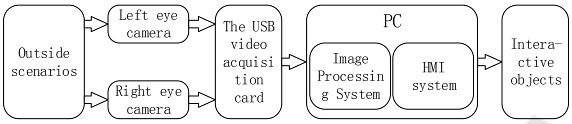


Figure 2: Binocular system diagram.

The head of the robot vision system is of two micro cameras installed in the eye model, camera resolution is 1280╳960, refresh rate is 50f/s.

The video capture card is used to transform the analogue signals obtained by the vision sensor into digital image data. In order to facilitate the development of the system on the mobile platform, we chosen TC-U652 video capture card with USB interface.

3 VISUAL SYSTEM

In this paper, the visual system can realize face detection, face location, gender recognition system and facial expression recognition and representation, which provides a comprehensive information for human-robot interaction, as shown in Fig.3. This paper use OpenCV as main development platform, programming with C + + language. We also use the QtCreator3.0.0, Qt5.2.0 and MSVC2010

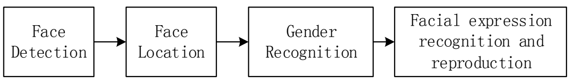


Figure 3: Visual system.

3.1 Face Detection

This paper adopt a face detection algorithm based on Haar features of the cascade classifier. The process is shown in Fig.4.

The operation will be initialized and the camera will be called by OpenCV to achieve image acquisition.

Then, the system will judge the status of camera. If the camera cannot be open, it will display error and exit the program. On the contrary, the binocular cameras open normally, each camera reads an image and determine whether it is empty, if the image data is empty, the system will display error and exit the program.

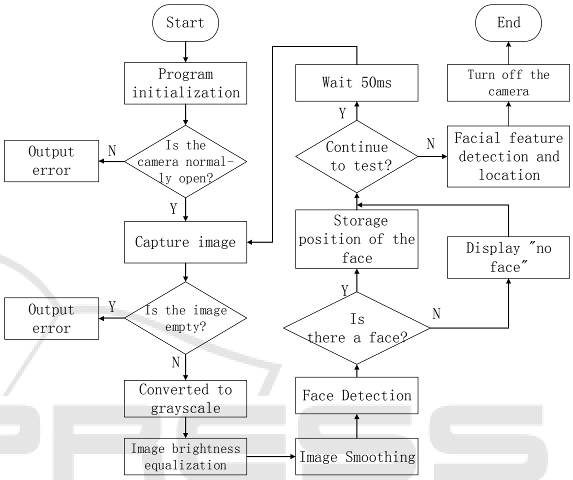


Figure 4: Face detection process.

After the image is read, the image need to be processed in order to improve the recognition rate of the classifier, including image grizzled, brightness equalization and smooth processing.

It use defined cascade classifier for face detection that based on Haar-like features (Peleshko and Soroka, 2013; Quanlong et al., 2014; Lienhart and Maydt, 2002) after the pre-treatment of the image. The image Detected is shown in Fig.5 (a).

Further processing is needed for practical applications. After realizing the face detection, we also design the program of the facial feature detection and location, including the eyes, nose and mouth, which have important significance to improve the accuracy of face detection. The trained classifier Haarcascade_eye.xml, Haarcascade_mcs_nose.xml and Haarcascade_mcs_mouth.xml are respectively detected eyes, nose and mouth, which will be marked with boxes of different sizes, as shown in Fig. 5 (b).

(a) Only detect human face

(b) Detect human face and face features such as eyes, nose and mouth

Figure 5: Face detection stage results.

The output of this paper is the variable size of the face image that can vary with the distance between face and the camera, which is different with the detected face images of the traditional output.

After the end of the test, it shows the recognition results. If we have the need, the system will get the next frame image and repeat the recognition process. This section must be in one column.

3.2 Facial Location

The face localization is the relative position between the robot and the target, and the relationship between the relative position and the parameters of the camera can be determined by establishing the binocular vision model. Internal parameters are obtained using MATLAB and self-calibration method.

Robot camera is the wide-angle lens. The focal length is small, which be generally within 30mm, and binocular vision system position in the larger scene. The subject and the distance between the camera is much larger than 30mm, that is, u >> f, so,

$$\frac{1}{v}=\frac{1}{f} \quad (1)$$

where u is object distance, v is image distance, f is focal distance.

Because of v≈f, the camera lens system can still use small aperture imaging model. Through the transformation between physical coordinates $O_c - X_C Y_c Z_c$ of image coordinate system and camera coordinate system $O_I - X_I Y_I$, the relationship can be established between the pixel (u, v) of image coordinate system and world coordinate system coordinates $O_W - X_W Y_W Z_W$, as follows:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M_1 M_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2)$$

where s represents Z_c, f_x and f_y are focal length in pixels, R is a rotation matrix, t is a translation vector, M is called the camera parameter matrix, M1 matrix is only related to the intrinsic parameters of the camera. It is the camera parameter matrix. The M2 matrix describes the position and orientation of the camera in the world coordinate system. It is the external parameter matrix.

The physical and geometric model of the robot binocular vision system is shown in Fig.6. The left and right eyes of the camera coordinate system use the optical centre as the origin respectively, the world coordinate system is located in the middle. Three coordinate system parallel to each other.

The focus length of left and right eyes are respectively (f_{x1}, f_{y1}) and (f_{x2}, f_{y2}), eyes spacing 2B. The optical centre coordinates are respectively (c_{u1}, c_{v1}) and (c_{u2}, c_{v2}), the distance between the two eyes is 2B.

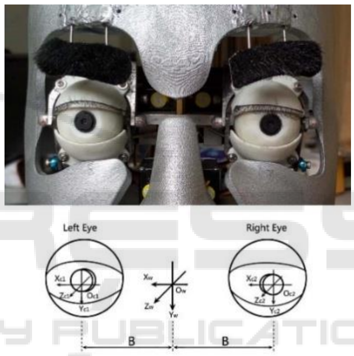


Figure 6: Binocular system and geometric model of facial robot.

The method of parallel to the optical axis orientation is used in this paper. The model of left and right eye camera and focal length are approximately the same, but the photo centric coordinates will make a difference due to assembly and manufacturing error. We can approximate that:

$$f_{y1} = f_{x1} = f_{y2} = f_{x2} = f \quad (3)$$

According to the above formula:

$$\begin{cases} z_w = \dfrac{2Bf}{(c_{u1} - c_{u2} - u_1 + u_2)} \\ x_w = z_w \cdot \dfrac{u_1 - c_{u1}}{f} + B \\ y_w = z_w \cdot \dfrac{v_1 - c_{v1}}{f} \end{cases} \quad (4)$$

The formula is the corresponding relationship of a little space point between the pixel coordinates and physical coordinates.

The traditional camera calibration method of grid calibration board is used in this paper, and the camera calibration is carried out by means of MATLAB toolbox. But because the camera calibration result is affected by the size of the calibration board, the size of the grid, the position of the display and different light, the results may be volatile. In camera calibration, compared with the focal length, the changes of parameter have greater influence to the optical centre coordinates. The overall image resolution is only 640 ∗480, the light fluctuation is too large and even unable to meet the needs of target positioning. We choose self-calibration method to calibrate the camera optical centre coordinates.

We came to the conclusion about the eye-parameter matrix, respectively:

$$M_{1L} = \begin{bmatrix} 938 & 0 & 381 & 418 \\ 0 & 938 & 218 & 224 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (5)$$

$$M_{1R} = \begin{bmatrix} 953 & 0 & 328 & 330 \\ 0 & 950 & 177 & 260 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (6)$$

3.3 Gender Identification

Gender classification is a typical two classification problem. The recognition process is divided into the following stages: image pre-processing, facial gender feature extraction and recognition analysis. Gender recognition based on face is mainly analysed by facial features, and then through the algorithm gender characteristics of the face are classified.

On the basis of face image pre-processing, gender recognition system is further designed, the program flow diagram as shown in Fig.7, mainly including PCA feature extraction and LDA classification.

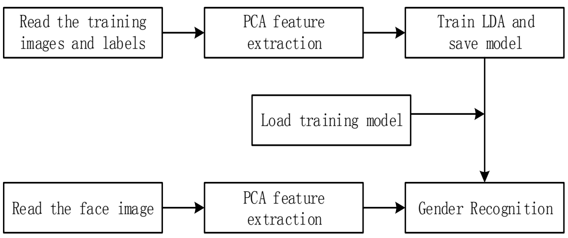


Figure 7: Gender Recognition System program flow chart.

In this paper, we choose the most widely used PCA to extract the gender characteristics of human face images, which extract the gender characteristics mainly, while reducing other information. The essence of PCA is to transform the original feature into the low latitude space in the case of the best possible representation of the source feature.

Then the classifier is designed by LDA for gender identification, which can extract the most discriminative low-dimensional feature from the high dimensional feature, and select the characteristics that reflect the maximum ratio of the samples between class dispersion and intra-class dispersion.

3.4 Facial Expression Recognition and Reproduction

The facial expression recognition and representation system is divided into four parts, including the pre-processing of the input image, feature extraction and dimensionality reduction of the processed image, the stage of facial expression recognition, the reproduction of facial expression recognition.

On the basis of image pre-treatment, the facial expression recognition system is further designed, and the flow chart is shown in Fig.8.

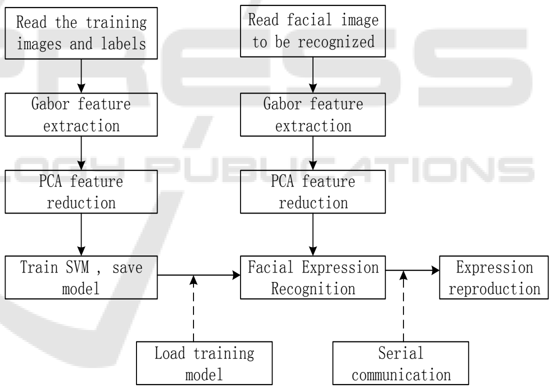


Figure 8: The system of facial expression recognition and reproduction.

In the system of facial expression recognition and representation, feature extraction is the key of the recognition technology. The effective method of extracting facial features will directly affect the level of recognition rate. In this paper, Gabor wavelet is used to extract facial expression features. Therefore, the case has been effectively improved with reducing Feature dimension feature reduction and preserving the important feature. For this purpose, PCA is used to reduce the dimension of the feature extracted by Gabor.

A classifier based on SVM is used for facial expression recognition, the method of SVM expression classification is the linear kernel function.

After identifying the expression, the PC sends instructions to the next bit machine through the serial communication and controls the robot to reproduce the expression.

4 EXPERIMENTAL EVALUATION

In this paper, the experiment is carried out on the platform of the vision system of humanoid head robot.

First of all, the accuracy of face location is verified, and the calibration line is positioned on the scale. The positioning results are compared with the calibration results and the positioning accuracy is measured. The vertical positioning experiment is shown in fig.9.

Figure 9: Vertical equidistant points Positioning Experiment.

The results of vertical equidistant points experimental are shown in Figure 10.

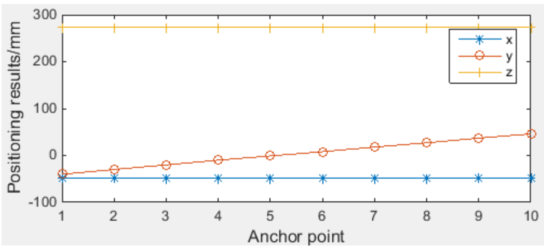


Figure 10: The results of vertical equidistant points.

From the figure, the results of X, Z axis positioning in ten measurement point are the same, the X axis distance has been -49mm, Z axis has been 273mm, which consistent with the actual value. The results of Y axis increase with the positioning point. The main error are on the Y axis and the positioning error are between 0-1mm, the error curve as shown in Fig.11.

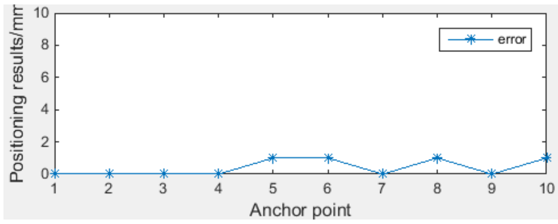


Figure 11: The error of vertical positioning

In the same way, the horizontal isometric point experiment is carried out. The main error is concentrated on the X axis that is in 0-3mm, and the error is between 0mm to 3mm.

Experiments show that the face localization can be successfully tested, and the accuracy and stability are great.

FERET (Phillips et al., 1998) face database is selected for the establishment of gender identification system, gender recognition experimental results shown in Figure 12. Experiments show that the robot can accurately identify the gender.

(a) Result: male (b) Result: female

Figure 12: Gender recognition experiments.

CK+ (Lucey et al., 2010) facial expression database is selected for the establishment of facial expression recognition and reproduction system. The experimental results of facial expression recognition and representation are shown in Figure 13.

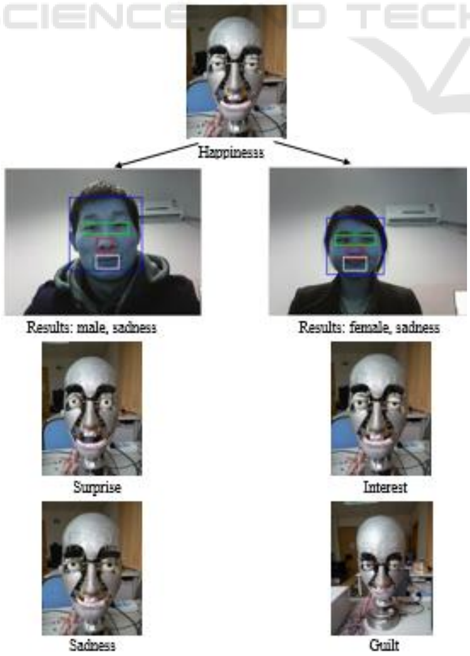
Apart from this, there are five other basic facial expression, which can show the realization of facial expression recognition and reproduction system. In addition to the six basic facial expressions, the system can also achieve the other four special expression, interest, arrogant, guilt and cachinnation.

(a) Identified surprised expression (b) Reproducing surprised expression

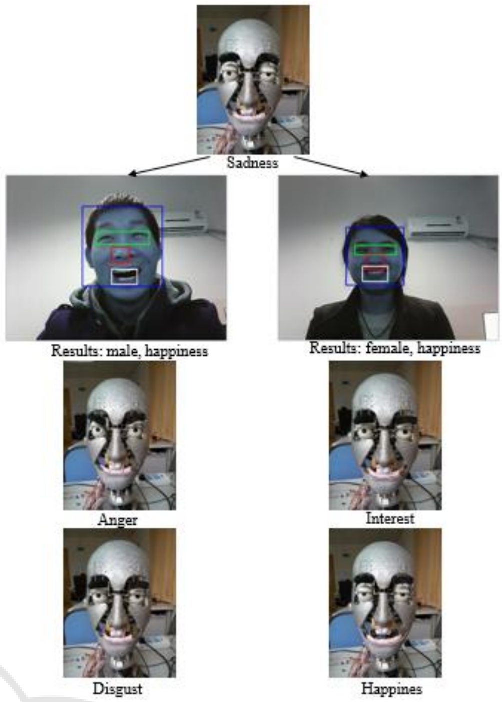
Figure 13: Expression recognition and reproduction experiments.

In terms of human emotional interaction, emotional information is obtained through the direct or indirect contact with human beings. Emotional information is analysed and identified by modelling. Perceptual understanding can be obtained by reasoning the results. Finally, the results can be expressed through the appropriate way, it completed the whole process of emotional interaction. The experiments are shown in fig.14.

Robot expression is happiness on face image recognition. If recognition is male and the expression is sad, the robot will become surprised, finally turn to sad; if the recognition is female and the expression is sad, the facial expression robot will become interested, then turn to guilty, such as shown in Fig.14.(a).



(a) The human-robot interaction process of the same expression for different gender when the initial expression is happiness



(b) The human-robot interaction process of the same expression for different gender when the initial expression is sadness

Figure 14: Human-robot interaction experiments.

Robot expression is sadness on face image recognition. If recognition is male and the expression is happy, the robot will become angry and finally turn to disgust; if the recognition is female and the expression is happy, the facial expression robot will become interested and eventually become happy, as shown in Fig.14 (b).

It can be seen in the figure that, for the same kind of expression of different gender, the robot's emotional change is not the same. The robot can make a harmonious human-robot interaction through vision.

5 CONCLUSIONS

This paper developed the vision system based on the SHFR- III facial expression robot, including face detection, face location, gender recognition and face recognition and representation, which are used for human-robot interaction. The results of experiments show that robot can successfully detect human face and locate face position accurately. It is able to identify 6 basic facial expressions and the machine can reproduce the expressions. Other special facial expressions can be achieved through the head-neck coordination. It also can correctly identify the gender of the recognized face image. By making a certain response to the detected face information, human-robot interaction can be achieved.

The future work is focused on the establishment of an emotional model for the facial robot SHFR-III, which is a robot with richer artificial intelligence.

ACKNOWLEDGEMENTS

This research is supported by National Key Scientific Instrument and Equipment Development Projects of China (2012YQ150087), National Natural Science Foundation of China (61273325).

REFERENCES

Delaunay F, Greeff D J, M. Belpaeme T, 2009, Towards retro-projected robot faces: An alternative to mechatronic and android faces, *The 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama*, pp. 306-311.

G. Trovato, T. Kishi, N. Endo et al, 2012, Development of facial expressions generator for emotion expressive humanoid robot, *2012 12th IEEE-RAS International Conference on Humanoid Robots*, pp.303-308.

Alberto Parmiggiani, Giorgio Metta, Nikos Tsagarakis, 2012, The mechatronic design of the new legs of the iCub robot, *IEEE-RAS International Conference on Humanoid Robots,Japan*, pp.481-486.

Peleshko D,Soroka K, 2013, Research of Usage of Haar-like Features and AdaBoost Algorithm in Viola-Jones Method of Object Detection, I*nternational Conference on the Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, pp.284-286.

Quanlong Li, Qing Yang, Shaoen Wu, 2014, Multi-bit sensing based target localization (MSTL) algorithm in wireless sensor networks, *2014 23rd International Conference on Computer Communication and Networks (ICCCN). Shanghai*. pp.1-7.

Lienhart R, Maydt J, 2002, An extended set of Haar-like features for rapid object detection, *2002 International Conference on Image Processing,* pp.900-903.

Yunsheng Han, Mobile Robot Positioning Target Object Based on Binocular, Jiangnan University.

Ekman P, Rosenberg E L, 2005, What the face reveals: basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)," *Second Edition. Stock: Oxford University Press.*

Lucey P, Cohn J F, Kanade T, et al, 2010, The extended Cohn-Kanade dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expression *Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis*, pp.94-101.

Phillips P J, Wechsler H, Huang J, et al, 1998,The FERET database and evaluation procedure for face recognition algorithm, *Image and Vision Computing*, pp.295-306.