Visually Gesture Recognition for an Interactive Robot Grasping Application

Kun Qian[1] and Chunhua Hu[2]

[1]*School of Automation, Southeast University, No.2 Sipailou, Nanjing, 210096, China*
[2]*School of Information Science and Technology, Nanjing Forestry University*
No.159 Longpan Road, Nanjing, 210037, China
kqian@seu.edu.cn

Abstract

Gesture based natural human-robot interaction paradigm has few physical requirements, and thus can be deployed in many restrictive and challenging environments. In this paper, we propose a robot vision based approach to recognizing intentional arm-pointing gestures of human for an object grasping application. To overcome the limitation of robot onboard vision quality and background cluttering in natural indoor environment, a multi-cue human detection method is proposed. Human body is detected and verified by merging appearance and color features with robust head-shoulder based shape matching for reducing the false detection rate. Then intentional dynamic arm-pointing gestures of a person are identified using Dynamic Time Warping (DTW) technique, whilst unconscious motions of arm and head are rejected. Implementation of a gesture-guided robot grasping task in an indoor environment is given to demonstrate this approach, in which a fast and reliable recognition of pointing gesture recognition is achieved.

Keywords: *Human Detection, Gesture Recognition, Vision-Based Human-Robot Interaction, Mobile robot*

1. Introduction

One goal of developing intelligent and interactive robots is to make them ware of the user's presence, comprehend his gesture instructions and react accordingly to complete the corresponding task. To perform a natural and intuitive interface between robots and non-instructed users, gestures of arm, face and hand are the most commonly used as a mean of non-verbal communication.

Recently, many studies [1-4] have been focused on recognizing arm-pointing gestures as a human-robot interface. These approaches depend either on surrounding sensors like a stereo vision system [5], immersive environment [6], or 3D tracking of face and hands [7]. All these approaches are not practical with the limitation of the robot's computational and perceptional abilities and the lack of surrounding sensors. And thus the working conditions for their system are very restrictive.

In this paper we present an approach for adapting gestures as a communication scheme in the Human-Robot Interaction (HRI) context. More specifically, we propose a practical method to recognize frontal arm-pointing gesture in monocular images that work with a moving PTZ camera, low image quality, complex backgrounds and changing lightening conditions. Measures of body verification and dynamic motion feature matching have been taken to improve the reliability of pointing gesture recognition and ignore other unconscious

gestures. 2D appearance features of arm angle and head orientation are utilized and thus computational efficiency is increased. Based on this method, implementation of a robot grasping application is developed.

2. Human Body Detection and Verification

Figure 1 shows our system framework. First, the detection module detects human body color (clothes) and discriminates it from other confusing objects by head-shoulder shape verifications. Then the recognition module recognizes the arm-pointing gesture in consecutive frames. Finally, the robot's action is generated to fulfill the task, such as a grasping and object fetching. We use a multi-cue method combining color and head-shoulder shape for human detection.

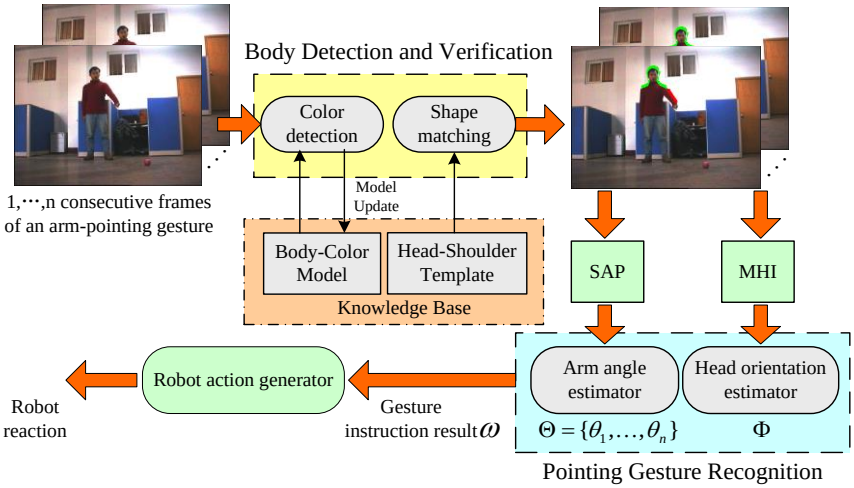


Figure 1. System overview

2.1. Robust Color Detection Using Color Probability Density Map

Firstly, a Gaussian color filter (Equation 1) is applied to each pixel in the each frame to build a color probability density map.

$$P(X_i) = e^{(X_i - \hat{X}_{body})^T \Sigma_{body}^{-1} (X_i - \hat{X}_{body})}, \quad (1)$$

where $X_i = [H_i, S_i]^T$ is the H and S value of the i th image pixel in HSI color space, $\hat{X}_{body}$ and Σ_{body} are the mean and covariance matrix of a body (clothes) color model. The filtered pixels are classified to binary value, and then the binary image is smoothed using a morphology closing operation. By applying a recursive searching algorithm and ignoring disturbing regions, the bounding box of body, $RECT_{body}$, is located in the image, which is taken as an observation of human hypothesis.

Then, to compensate the illumination changes, $\hat{X}_{body}$ and Σ_{body} of the Gaussian Model are adapted. In k th ($k > 1$) frame, the robot computes new mean $\hat{X}_{body}^{new}$ and covariance $\hat{\Sigma}_{body}^{new}$ from small rectangular regions around the center of bounding box located in $(k-1)$ th frame, and

the two parameters of Gaussian Model are adapted with a updating rate α, which is set to 0.1 in the experiment.

$$\hat{X}_{body}^{k} \leftarrow \alpha \hat{X}_{body}^{k-1} + (1-\alpha)\hat{X}_{body}^{new} \quad (2)$$

$$\hat{\Sigma}_{body}^{k} \leftarrow \alpha \hat{\Sigma}_{body}^{k-1} + (1-\alpha)\hat{\Sigma}_{body}^{new} \quad (3)$$

Figure 2 shows an example of color detection. In Figure 2(a), darker pixel indicates higher probability of the body-color; in Figure 2(b), the search window is denoted with a red rectangle, with the biological assumption that head is 1/3 the height of torso, and above the torso; in Figure 2(c), edges are extracted within the scope of the search windows.

(a) (b) (c)

Figure 2. Color Detection

2.2. Human Target Verifications

After the image pre-processing, a human verifications technique using head-shoulder shape matching in the distance image is employed to improve the accuracy of body detection. Head-shoulder contour is a prominent feature of upper-body shape with the underlying assumption that the head-shoulder is almost invariant during human motions. A Median Filter is applied to smooth the gray-value distribution, followed by edge extraction using Canny Detector within certain search window (shown in Figure 2(c)). Then we use an improved Euclidean Distance Transforms (EDT) method, which calculates the Euclidean Distance of certain neighbors around the edge pixel, rather than scanning the whole image. The method generates the Edge Distance Image (EDI), which is more accurate and much less time-consuming than conventional Chamfer Matching [8] algorithm. Finally, an edge matching scheme using Hausdorff distance is conducted.

3. Pointing Posture Recognition

1) Left or Right Arm Raised

The horizontal gravity center C_{x1} and the centroid-of-area C_{x2} are computed and compared within the area of $RECT_{body}$. *If* $(C_{x1} > C_{x2})$, the gesture $\omega \in$ *Left-Arm-Pointing,* and otherwise , $\omega \in$ *Right-Arm-Pointing*.

2) Pointed arm angle estimation

The upper-half contour of the raised arm is labeled and Least-Squares (LS) Line Fitting method is employed to estimate the fitting linear equation: $y = k \cdot x + b$, and $\theta = c\tan^{-1}(k)$, which measures the angle between the raised arm and vertical axis. Figure 3 depicts an example of arm angle estimation.

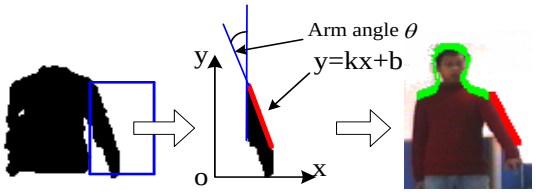


Figure 3. Arm angle estimation. (a)ROI of arm (b) Arm fitting line (c) head-shoulder fitting

3) Dynamic arm angle matching using DTW

When the pointing gesture is performed, arm angle sequence $\Theta = \{\theta_1, \ldots, \theta_n\}$ is estimated in consecutive frames and recorded into a profile called Sequential Angle Profile (SAP). By analyzing the SAP map of different arm motion gestures, it reveals that the SAP serves as a sufficient description of the pointing arm gesture with time information. Figure 4 illustrates the Motion History Images (MHIs) of three typical gestures.

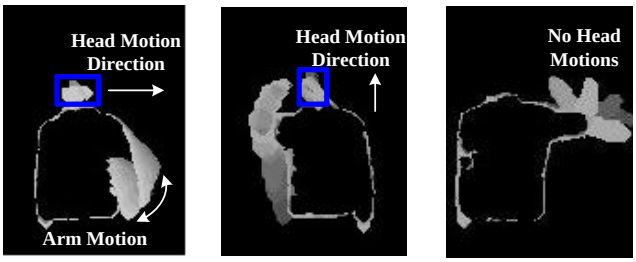


Figure 4. The Motion History Image (a)A typical pointing gesture. (b)An unconscious raising arm gesture (c)An non-pointing gesture

To handle the variations in temporal behavior, the matching is computed using Dynamic Time Warping (DTW) algorithm [9]. Figure 5 shows the Sequential Angle Profile (SAP) of three different gestures that corresponds to the three examples in Figure 4. As shown in Figure 5, the SAP of three arm gestures varies significantly. Essentially, the DTW algorithm finds the "best fitting" correlation *Dist* between the angle sequences of testing frames and motion template by searching along the "best fitting" path. Thus, by thesholding *Dist* , a judgment can be made whether the gesture is a pointing one.

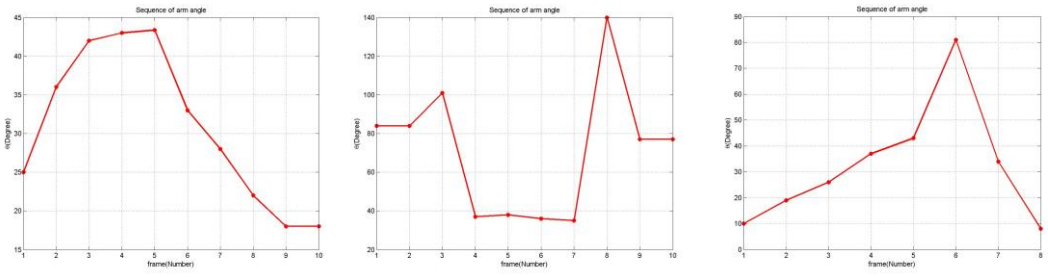


Figure 5. The SAP of three example gestures

4) Head orientation estimation

Besides the dynamic matching of arm angles, head motion is also utilized as additional information of an intentional pointing. With the difficulty in estimating the gaze direction in every frame, head orientation Φ is also estimated by analyzing the head motions during the gesturing time. A significant horizontal change of head in the MHI indicates a general horizontal movement of head, which is taken as a symbol of *Gazing-At-Hand*.

4. Experiments and result

This approach is tested on ActivMedia PIONEER -2DX mobile robot with different lightening conditions and complex backgrounds. Figure 6 demonstrates the configuration of the experiment. A user stands γ meters in front of the robot, raises and points at a position (θ,φ) of a ball on the ground, while gazes at the pointed hand, where $\theta \in [0^\circ, 90^\circ]$, $\gamma \in [2\text{m}, 3\text{m}]$, $\varphi \in [60^\circ, 120^\circ] \bigcup [-120^\circ, -60^\circ]$. Furthermore, a gesture-based human-robot interaction is implemented for a ball grasping application.

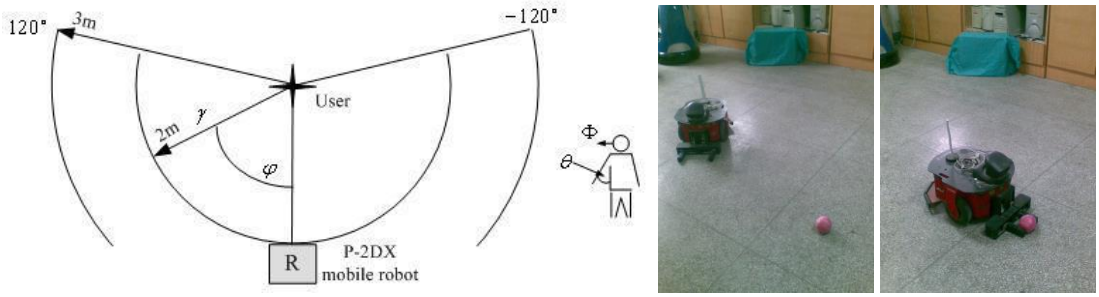


Figure 6. The configuration and the pointing-directed grasping application

Figure 7 elucidates the results. If the likelihood to human shape falls above *Threshold* , which is set to 80% in the experiment, it is regarded as frontal human body, and his raised arm is fitted. Compared with the Adaboost [9] method proposed by Viola and Jones, our method has the advantages in that: it is capable of tracking a side view or rear view body and the false detection rate can be lowered, as a comparison shown in Figure 7(e), which indicates that the Adaboost algorithm is prone to false detection of human.

(a) (b) (c) (d) (e)

Figure 7. Likelihood to human shape and arm fitting

We built a motion template of sequential angle from a typical pointing gesture: $\Theta_t = \{0,10,20,30,\ 40,50,40,30,20,10,0\}$. Table 1 illustrates the recognition result of the three gestures in Figure 4. Besides, we tested 20 pointing gestures of different speed with 18 samples identified correctly and 30 other different dynamic arm gestures with 28 samples

identified correctly. In the experiment, two continuous frames with the same arm angle is taken as the start and end of the gesture.

Table 1. Recognition result

Gestures	Sequential Angles	*Dist*	Recognition Result
Fig.4(a)	{0,25,36,42,43,38,33,28,22,18,18,0}	9.7	*Left-Arm-Pointing*
Fig.4(b)	{0,10,19,26,37,43,81,34,8}	14.4	*Non-Pointing*
Fig.4(c)	{84,101,37,38,36,35,140,77}	47.4	*Non-Pointing*

4. Conclusion

The uniqueness of the proposed method is that it is a low-cost oriented approach to recognize arm-pointing gestures only using monocular images of relatively poor quality. The experimental results show that, the body verification rejects non-users, and the dynamic gesture recognition combining sequential arm angles and head orientation ignores unconscious motions. Hence the reliability of the system is increased, meanwhile computational efficiency is maintained. An interactive robot grasping task is implemented, which indicates that the system work fast and reliable on a mobile robot with limited computing and sensing abilities. In the future work, we are interested in utilizing a distributed sensor network as a supplement of robot's abilities and investigate the new human-robot interaction problems in such an intelligent environment.

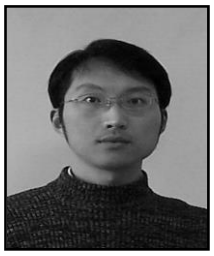
Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 61105094) and the open fund of Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education (No. MCCSE2012B02).

References

[1] A. Xu, G. Dudek and J. Sattar, "A Natural Gesture Interface for Operating Robotic Systems", Proceedings of the IEEE International Conference on Robotics and Automation, Pasadena, USA, **(2008)**, pp. 3557-3563.
[2] N. Nguyen-Duc-Thanh; S. Lee and D. Kim, "Two-stage Hidden Markov Model in gesture recognition for human robot interaction", International Journal of Advanced Robotic Systems, vol. 9, no. 39, **(2012)**.
[3] H. H. Kim, Y. S. Ha, Z. Bien and K. H. Park, "Gesture encoding and reproduction for human-robot interaction in text-to-gesture systems", Industrial Robot, vol. 39, no. 6, **(2012)**, pp. 551-563.
[4] M. H. Ju and H. B. Kang, "Emotional interaction with a robot using facial expressions, face pose and hand gestures", International Journal of Advanced Robotic Systems, vol. 9, no. 95, **(2012)**.
[5] Y. Yu, Y. Ikushi and S. Katsuhiko, "Arm-pointing Gesture Interface Using Surrounded Stereo Cameras System", Proceeding of the 17th International Conference on Pattern Recognition, Cambridge, England, UK, **(2004)** August 26-26, pp. 965-970.
[6] K. Roland and V. G. Luc, "Real-Time Pointing Gesture Recognition for an Immersive Environment", Proceedings of the 6th IEEE AFGR, Seoul, Korea, **(2004)** May 19, pp. 577-582.
[7] K. Nickel, E. Seemann and R. Stiefelhagen, "3D-Tracking of Head and Hands for Pointing Gesture Recognition in a Human-Robot Interaction Scenario", Proceeding of the 6th IEEE AFGR, Seoul, Korea, **(2004)** May 19, pp. 565-570.
[8] G. Borgefors, "Hierarchical chamfer matching: a parametric edge matching algorithm", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 2, **(1998)**, pp. 849-865.
[9] A. D. Wilson and A. F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition", IEEE Transaction on Pattern Analysis and machine intelligence, vol. 21, no. 9, **(1999)**, pp. 885-899.
[10] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", Proceedings of the International Conference on Pattern Recognition, **(2001)**, pp.515-518.

Authors



Dr. Kun Qian received the Ph.D. in control theory and control engineering from Southeast University in 2010. He is currently working as a lecturer at the School of Automation, Southeast University. His research interests are intelligent robotic system and robot control.

Associate prof. Chunhua Hu received the Ph.D. in control theory and control engineering from Southeast University in 2008. She works in the School of Information Science and Technology, Nanjing Forestry University since 2012. Her research interests are robot vision and pattern recognition.