Design and Implementation of User Centered Interface on Mobile Phone

Hoa-minh LE, Kwang-jin HONG, Kee-chul JUNG

(*Department of Media*, *Soongsil University*, *Seoul* 156-743, *Korea*)

Abstract – **Mobile device research has been increasing rapidly in Human Computer Interaction (HCI) recently. Following this trend, this paper proposes a user-centered interface, which has been designed, completely installed and independently run on a mobile phone. Video signal is streamed through its camera as the image input to the interface by employing techniques of image processing, computer vision and graphics to identify automatically absolute positions of human face, neck and two hands. A paradigm is also put up theoretically. And it embeds this interface to perceive the human postures and convert them into relaxed comic character according to its context.**

Key words – *user-centered interface*; *human postures*; *comic character*

Manuscript Number: 1674-8042(2010)04-0330-04

dio: 10.3969/j.issn.1674-8042.2010.04.07

1 Introduction

The explosive growth of mobile device has yet to be fully exploited by HCI researchers[1]. It is still a gold mine for scientists to discover its potential. Equipped with powerful chips, mobile phone now can load computing capabilities from computer and await researchers skilled in HCI to create new tools[2], which ease everyday tasks or simply joyful entertainment since it gradually becomes an inseparable thing. Our proposal including designing and also implementing an interface to contribute a bridge for interactive systems such as human machine interaction or user oriented mobile applications is not outside this track.

Computer science techniques are utilized behind the stage in this interface. Various algorithms in facial recognition had been compared and successfully implemented to detect human face in a 2D image[3]. Whereas human skin detection[4-5] and K-means clustering[6] had been used to classify different parts of a human body. Taking the full advantage of graphics technique, we refined the image processing steps including conversion between color spaces YUV, RGB[7-9], downsizing or gray scaling to optimize them for minimum computation cost.

As opposed to other previous related researches due to expensive processing on mobile phone, which just was a blue print[10], stopped in the detecting colored objects to control humanoid soccer robots[11], mobile phone was either used to strict human area constraint[12] or played as a human-robot speech recognition interface recently[13].

The work is novel as a modality which is capable of pointing out human body parts′ position (Fig. 1) and running independently on an Android phone, specifically on Google Nexus One[14]. Moreover it extends potential to be the default physical interface for ubiquitous computing applications[15]. Inspired by this interface, it is feasible for us to assume an experiment that supports for analyzing and process the user′s postures in the real world and translate them into humorous computer-generated graphical objects which are comic characters selected from the database. Furthermore, we can add daily funny dialogue into the character′s talk bubbles regarding to the context.

Fig. 1 Interface was implemented on Google Nexus One

※ **Received:** 2010-09-13

Project supported: This work 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-2009-(C1090-0902-0007)) and the Soongsil University BK 21 Digital Media Division

Corresponding author: Kee-chul JUNG(kcjung@ssu.ac.kr)

2 Design and implementation

Recently, mobile phone is not considered as a purely traditional telecommunication device, but it also represents a portable entertainment user oriented device and an efficient office assistant. Thanks to dramatically advance in hardware, now we can compress many techniques which acquire expensive computation such as image processing, computer vision and graphics into an interface placed totally on an Android phone, Google Nexus One.

Main components are video frames capturer, human skin color segmentation, body regions classification, face detection preparation, face detector and global position calculator (Fig. 2).

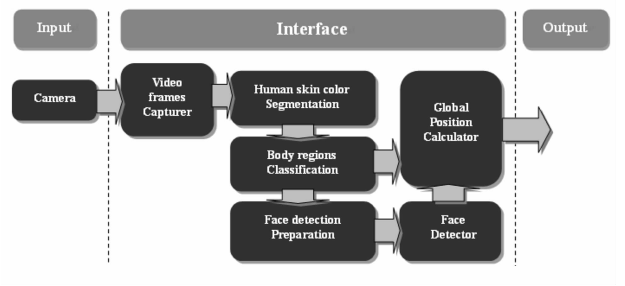

Fig. 2 Interface function blocks diagram

Streaming video frames through nexus one's camera acts like an input of the interface, are processed by video frames capturer. It also has a sub function is conversion between different color spaces when upcoming video stream is compressed in YUV format. Human skin color segmentation filters human skin regions from every frame. And then body regions classification will group skin regions together using K-means clustering algorithm. If area of any regions of interest is smaller than a specific threshold, it will be considered noise and be eliminated. It aims to de-noise when noise is small area and discrete pixels which existing in image frame. Throughout this step, the ideal result is the center positions of human face, and two hands. Face detection preparation has a task of preparing images for Face Detector. It downsizes the image and converts it to grayscale to increase processing speed. Human face region is detected by Face Detector and its position is identified accurately in the image frame.

Consequently, global position calculator computes and outputs precisely positions of user's face, neck, two hands. Clearly it has to combine two returned results from two threads. Comparing with face detector, it will know the face region. Neck region is right below and stick to the face area. Hence there are only two hands' areas left. Now the interface's output is ready for various applications based on human gestures.

3 Main algorithms

3.1 Human skin color segmentation

In this part we applied a constructive induction approach to find a set of new attributes for skin detection[4-5]. It starts with the three basic color components RGB in a normalized form and a simple set of arithmetic operators to produce a suitable model for skin detection. Once a new set of attributes is produced, a new restricted covering algorithm, called RCA, is used to construct a single rule of no more than a small number of easy to evaluate terms with a minimum accuracy. We are interested in inducing simple models as they are relevant to applications which require fast response times.

The rule for the first color model is: If $\frac{r\times b}{(r+g+b)^2}>0.1$, and $\frac{r\times g}{(r+g+b)^2}>0.1$ then Class = skin, with 93.7% of recall, 91.7% of precision and 92.6% of success rate. In order to increase the performance values of RCA, it is allowed to grow an additional term. The best model has the following performance results: 94.1% of recall, 92.7% of precision and 93.4% of success rate.

3.2 Body regions classification

This function uses K-means[6] at minimizing an objective function, in this case a squared error function. The objective function

$$j = \sum_{j=1}^{k} \sum_{n\in S_j} \left| x_n - C, \right.$$

where $|x_n - c$ is a chosen distance measure between a data point and the cluster centre c_j .

Although it can be proved that the procedure will always terminate, the K-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The K-means algorithm can be run multiple times to reduce this effect.

K-means clustering algorithm enables to group adjacent skin color points into clusters and find out the means. It eliminates clusters that are small. By this way noise also is eliminated automatically.

3.3 Face detector

Face detector is generated based on Adaboost algorithm. Adaboost is a kind of method to construct a strong classifier from a lot of weak classifiers. The weak classifier for face detection applies Haar-like features. This feature value is different in terms of intensity between pixels in particular location. One weak classifier is comprised of feature threshold, parity and function to compute feature value. Based on this concept, initially tens of thousands weak classifiers are created.

The first weak classifiers are trained by using Adaboost algorithm. Learning process consists of 4 steps as follows: ① determining feature threshold and parity through applying all weak classifiers to training images; ② computing error rate of each feature; ③ collecting weak classifiers which have low error rate; ④ generating a strong classifier using collected weak classifiers. For more details[3], explains the Adaboost algorithm.

The performance of face classifier based on Adaboost algorithm is already verified in many previous researches and the face detection rate is generally over 90%. Therefore Adaboost algorithm based face detector is the most widely used method in pattern recognition field.

4 Experiment

In this section, we present an empirical evaluation of our proposed interface on system performance and future work.

4.1 Runtime result

After calculating inside and outside with image frame size is 480×320 pixels (Fig. 1), we got the average result below, time period is measured by millisecond.

Tab. 1 Skin and face detection time

	Skin	Face	Total	Frame rate(f/s)
Inside	30	439	469	2.13
Outside	41	442	483	2.07

Consequently, the system processes approximately 2 frames/second. In the Tab. 1, it takes a little more time to detect human skin outside when the light and background texture changes. However there is a slightly difference with face detection phase. In fact, illumination does not influence the face detection algorithm too much at all, while it is a real pain in the neck as it makes a significant overlap between the skin and non-skin areas. For this reason, background and clothes color should be selected distinguish from the human skin color to avoid false positive.

In Ref. [14], the convergence of computing capabilities onto the mobile phone is predicted and we want to check it out on Nexus One. 2 frames/sec is still not suitable for real time human gesture based applications; however, it proves that the trend is not fictional, mobile devices proliferated even further in society in a greater variety of contexts than our current user models accommodate.

4.2 Further discussion

Entertainment has contributed a lot to the society. Entertainment basically fights boredom and gives us a temporary escape from certain worries. It changes the way we view the world in the positive way. We assume a use case which recognizes human gestures in the Augmented Reality Entertainment System through our proposed interface. Picture yourself walking down the street; you make fun of a friend or personal photos with different comic character depending on user posture. For example in Fig. 3, a character from The Simpsons will replace a talking friend and he/she is going to burst out laughing when see it.

This is a demonstration showing how to implement our proposed interface to recognize human gestures in the Augmented Reality Entertainment System (Fig. 4). Character Selector is mainly in charge of selecting the most suitable comic character from database that is a lot look like the input human posture. In a nutshell it receives parameters from our proposed interface and separates user posture into four areas base on his/her face and neck position. After that it calculates two angles of two hands against to the axis and relative distances from two hands against the neck to find out a nearest pattern from the comic characters database. It is easy to calculate these parameters when we got the accurate positions of face, neck and two hands from the interface. Moreover they have to be added approximate error to make sure that the system still can be tolerable with different trials.

Fig. 3 A comic character from the simpsons (™© FOX)

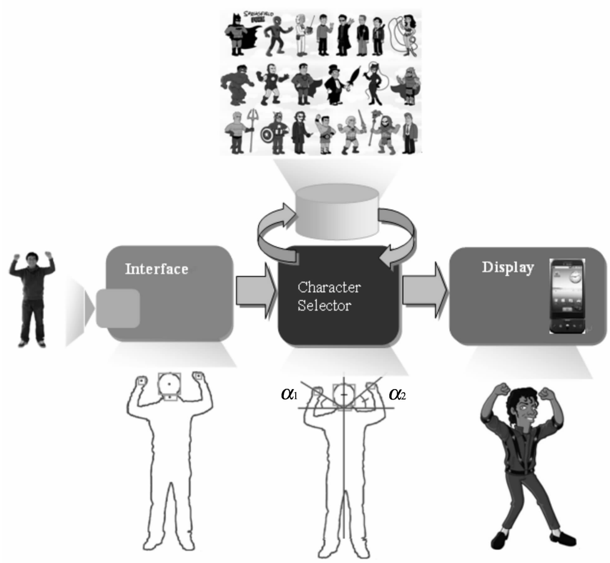

Fig. 4 Augmented Reality Entertainment System Paradigm (Comic characters ©: http://springfieldpunx.blogspot.com)

5 Conclusions

Recognizing that the extension of processing power into a mobile phone would become a normal phenomenon when there are many breakthroughs in hardware technology. Not only is the proposed ioterface inherited by those but also takes advantages from computer science technology in order to run independently on a mobile phone.

It is ready for human oriented applications which need precise position parameters of human body parts as a control bridge such as human-machine interaction, ubiquitous computing or mobile game. Our use case is just an open suggestion.

References

[1] J.O. Wobbrock, 2006. The future of mobile device research in HCI. Proc. of the workshop on "What is the Next Generation of Human-Computer Interaction?" at CHI 2006 Montreal, Canada, p.131-134.

[2] K. Y. Huang, 2009. Challenges in Human-Computer Interaction Design for Mobile Devices,. Proc. of the World Congress on Engineering and Computer Science WCECS, San Francisco, USA.

[3] P. Viola, M. Jones, 2004. Robust real-time face detection. *International Journal of Computer Vision*, 57(2): 137-154.

[4] G. Gomez, E. Morales, 2002. Automatic Feature Construction and a Simple Rule Induction Algorithm for Skin Detection. Proc. of Workshop on Machine Learning in Computer Vision, p.31-38.

[5] P. Kakumanu, S. Makrogiannis, N. Bourbakis, 2007. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40(3): 1106-1122.

[6] J. B. MacQueen, 1967. Some Methods for classification and Analysis of Multivariate Observations. Proc. of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, p.281-297.

[7] M. Sáenz, P. Salama, K. Shen, E. J. Delp, 1999. An Evaluation of Color Embedded Wavelet Image Compression Techniques. Proc. of the SPIE/IS&T Conference on Visual Communications and Image Processing (VCIP), San Jose, California, p.282-293.

[8] K. Jack, 1996. Video Demystified. 2nd ed., HighText.

[9] C. A. Poynton, 1996. A Technical Introduction to Digital Video. John Wiley & Sons.

[10] H. Huttenrauch, M. Norman, 2001. PocketCERO-mobile interfaces for service robots. Proc. of MobileHCI'01, IHM-HCI, Lille, France.

[11] S. Behnke, J. Stückler, H. Strasdat, M. Schreiber, 2007. Hierarchical Reactive Control for Soccer Playing Humanoid Robots,. M. Hackel (editor): Humanoid Robots: Human-like Machines, I-Tech, p.625-642.

[12] S. Calinon, A. Billard, 2003. PDA Interface for Humanoid Robots,. Proc. of the IEEE International Conference on Humanoid Robots (Humanoids), Karlsruhe, Germany.

[13] W. Kacalak, M. Majewski, 2009. Natural Language Human-Robot Interface Using Evolvable Fuzzy Neural Networks for Mobile Technology,. ICIC 2009, LNCS 5754, Springer-Verlag Berlin Heidelberg, p.480-489.

[14] R. Meier, 2009. Professional Android? Application Development. Wiley Publishing, Inc.

[15] R. Ballagas, J. Borchers, M. Rohs, J. G. Sheridan, 2006. The smart phone: A ubiquitous input device. *IEEE Pervasive Computing*, 5(1): 70-77.