Scientific
Research

Design and Implementation of the Image Interactive System Based on Human-Computer Interaction

Shanwei Song[1], Yuting Wu[2], Fang Zhang[2]
[1]*School of Software Engineering, Shanghai Jiaotong University, Shanghai, China*
[2]*School of Software Engineering, Tongji University, Shanghai, China*
E-mail: *jane.wu@whu.edu*

Abstract

Based on the traditional Human-Computer Interaction method which is mainly touch input system, the way of capturing the movement of people by using cameras is proposed. This is a convenient technique which can provide users more experience. In the article, a new way of detecting moving things is given on the basis of development of the image processing technique. The system architecture decides that the communication should be used between two different applications. After considered, named pipe is selected from many ways of communication to make sure that video is keeping in step with the movement from the analysis of the people moving. According to a large amount of data and principal knowledge, thinking of the need of actual project, a detailed system design and realization is finished. The system consists of three important modules: detecting of the people's movement, information transition between applications and video showing in step with people's movement. The article introduces the idea of each module and technique.

1. Introduction

As we all know, Human-Computer Interaction, New Media Arts, Image Processing and Patten Recognition are very popular research subjects and directions currently.

A Human-Computer Interaction program will be achieved which will display images through Image Processing and Pattern Recognition. This program will bring you a new experience because of the new and unique camera-based capture of non-contact input mode. At the same time, it is also an experimental method which can separate objects from images and more complete detection of object motion by a collection of various images processing algorithms. What's more, this system is designed as a prototype of exhibition items and will help the future development of the items become more maturity, more variation.

2. Design of the Image System Based on Human-Computer Interaction

2.1. System Architecture Overview

The Image Display System based on Hume-Computer

Interaction is designed for the items which showed in the exhibition and has multi-media interactive features in one. Therefore, the hardware components are diverse and the diversity of the hardware has promoted the complexity of the software.

In the hardware system, it includes camera (or similar image capture devices), image processing computer and projector (or similar image display devices).

And for the software system, it consists of two parts, one is image processing application, and the other is the image display application. These two parts communicate through the pipeline.

We will introduce the hardware system and the software system respectively.

2.2. Hardware System Architecture

According to the demand of the whole system, the hardware system is consisting of camera (or similar image capture devices), image processing computer and projector (or similar image display devices).

In the usual experiments, we can use the low end of normal plug-in camera for testing and research.

In actual exhibition, it should use the industrial high-

speed camera instead. The industrial high-speed require the image processing computer must equip with the Gigabit Ethernet Network Interface Card so that the small local area network can connect with the camera. Meanwhile, the image processing computer's configuration should have a corresponding increase to meet the high-speed frame capture.

Currently, the home ordinary camera's catch rate is about from 7 frames per sec to 25 frames per sec. In the experiment, you can use 25 frames per second plug-in camera.

In actual exhibition hall, we need to configure the camera for industrial whose capture speed can be up to 70 frames/second to ensure the smooth of image capture. In addition, in order to improve the accuracy of the finally locate cordinates, it need to make the corresponding computer's calculation speed be up to 30 frames per second in capture and processing the number of images under the condition that processing two applications simultaneously.

In the arrangement of the official exhibition hall, the requirement of the computer's CPU is above 2.0 GHz dual-core (including 2.0). At the same time, respond to the image display application's needs, more than 1G memory are required (including 1G). Display Card is asked more than 128 m video memory.

Of course, projector should support a minimum of 1024×768 display resolution, 3000 lumens or more. HD is better to ensure the display effects of the video image.

2.3. Software System Architecture

According to the design of the hardware architecture in the whole system, software system is divided into two parts: the camera image processing applications and the image display application.

Camera image processing program obtains images from the camera and do a series of processing, so that to capture human motion and displacement information; Image display program is user-oriented, this application complete the movement and displacement corresponding to the video response with users. Between the two applications at the same time, it will have some means of communication to connected to ensure the coordination.

3. Design and Implementation of Modules

According to the demands of the entire system and the design of the hardware architecture and software architecture, this system will complete the design and implementation of the three main modules.

1) Human motion detection module;
2) Data signal transmission module;
3) Image processing and response module.

3.1. Human Motion Detection Module

The functions of the Human motion detection module are:

1) Capture people's movement image information from the camera.

2) Processing the images which captured, then analyzed them to get the displacement of people and speed information.

Implementation of this module is the key to the system to achieving human-computer interaction. First, system needs to capture frames from the camera. The speed depends on the camera and the computer's performance. The more images captured per second, the more accurate and real-time in people's movement and displacement calculation. It will have a great influence the on the overall performance of the system.

The main algorithm used here for motion detection is frame difference.

The entire process is as follows: For the captured images, it need to do pre-poor between this image and the background image beforehand first of all, and then binary this image to determine the region of interest ROI (Region of Interest); For the binaried image, it needs a series of image pre-processing, including filtering, morphological transformation and so on to reduce noise impact on ROI; Last, analyzed the ROI to reached its mid-point coordinates and other information, and this mid-point can be regard as the information of a person in the course of the movement's current location; From the adjacent two of the location information and the time difference, we can obtain the velocity.

3.2. Data Signal Transmission Module

The function of Data signal transmission module is to achieve the transmission of the human motion detection information and to ensure the completeness of the necessary information when the final image displayed. Image capture processing program (A-side) creates the named pipes. Since then, the image display program (B-side) will be using the same channel name for data extraction.

In the A-side, in order to guarantee the image processing calculation process and enter information to the pipeline can be achieved in parallel, we can use multi-threading technology.

Parent thread creates pipeline, and does the image processing. Between the parent threads and the child threads, we use the signal technology to ensure collaboration between them. When the parent threads get one displacement information after processing the image, the child thread will input the motion and displacement information which be kept in the public to the pipe.

In the B-side, as the image display program using the

Director MX 2004, during the communication between processes, in order to enable the smooth transfer of data, we need to use the SDK provided Director MX 2004 to write Xtras (plug), and receive data from the sink of pipe. Xtras is a foundation based on COM (Component Object Model) object for the model development and is the externsion of Director.

3.3. Image Processing and Response Module

The function of Image processing and response module is to control the final display of image video, it needs to correspond with the movement and displacement captured from the camera.

This module will achieve the total control of video display images, including the constant speed forward, fast forward, normal speed back, fast back, jump, back to the video head and so on. These different video operations are consistent with the motion and displacement results analyzed by computer.

In a certain catching range of the camera, we calculate the degree of displacement in a certain line. Suppose the size of the image captured by the camera is 640×480 (in general the size of image depends on the resolution of camera), it is desirable in the displacement of the image along the x-axis, then the range of activity is 0-640 (based on region of interest setting, this can be any part from 0 to 640).

Similarly, for display video images, assumed to be 60 seconds of video. When we get the speed and the proportion coordinates calculated by image processing module(x coordinate/Horizontal size of the image), set the start of video playback to be the proportion of the coordinates of people multiplied by the length of the video, the same as the playback speed conversion according to certain proportion rate of people. When people stop, the image stops.

Director MX 2004 support avi, rm, mov, Flv format. However, current version only supports partial control, such as pause, play, and stop for AVI. Only for MOV format we can achieve complete control, including play, pause, stop at the starting point, continue to play, fast forward, rewind, enter and pause, back and pause.

Therefore, for the final video played in system, we choose MOV. For other formats such as AVI, MPEG, etc., We can use SUPER (Simplified Universal Player Encoder & Renderer) or other video format conversion software to convert.

4. Experimental Analysis

The experimental video length is 2949, under the condition that the movierate is 1, each frame play 3, and the speed of director playback is 25 frames/sec. Therefore, when movierate is 1 we need 39.4 seconds to complete the entire playing process. According to statistics, the average speed of normal person walking is about 1.4 m/sec. While the catching range of the camera (even industrial camera) will not capture more than 5 meters, because that too long distance will affect the image quality. But if we use multiple cameras to capture theoretically speaking, it will be no limit to the size range. So we assumed that the capture range is about 4 meters width. The result is we need 3 seconds in normal pace.

Thus, under this condition, the video length is 2949, the range of linear movement is 4 meters, camera resolution is 640×480, and the capture speed is 25 frames/sec. The relations between basic displacement and movierate are shown in **Tables 1** and **2**.

We can infer from **Tables 1** and **2**, Motion and displacement and the Movierate's corresponds are determined by a certain degree of movement. As in real cases, camera can not capture every tiny movement, we modified above data according to the testing of director.

Since the capture range and the accuracy of the camera are limited, it makes the test so long that when the final system is running, the content of image displayed is too little and often jump some parts.

Therefore, in the video that final displayed, it demands concise, so we recommend to use multiple cameras to increase the capture range.

Table 1. Displacement and the playback speed of movement.

Motion and Displacement (D)	Movierate
$-640 \leq D < -80$	20%
$-80 \leq D < -40$	16
$-40 \leq D < -20$	(8).
$-20 \leq D < -10$	6.
$-10 \leq D < -5$	4.

Table 2. Displacement and the playback speed of movement continued.

Motion and Displacement (D)	Movierate
$-5 \leq D < 0$	2.
$D = 0$	0
$0 < D < 5$	2
$5 \leq D < 10$	4
$10 \leq D < 20$	6
$20 \leq D < 40$	8
$40 \leq D < 80$	16
$80 \leq D \leq 640$	20

5. Summary and Outlook

In general, this article has research some point in traditional and being developed human-computer interaction, introducing a new way of human-computer interaction and the current status and future development of HCI. For the system designed it uses human motion capture as a way to interact, and do some analysis on principles and basic techniques.

Nowadays, many methods are used for moving object detection, for example:

1) Gradient edge detection method, because the edge is often the greatest changes in the gray areas of the entire image. According to this feature we can extract this from the object edges precisely.

2) Through the template matching, feature value. That is to say we record the characteristics of the object, and then do comparison with the objects of image by a certain degree of similarity.

All the methods above, only from one frame can we extract the object. We use frame difference method in this system, it can improve the utilization of frames. However, the relative calculation and image pre-processing demand more.

Secondly, all the camera image processing is based on a camera. This resulted in the capture range that the range become too narrow and to the disadvantage of capturing people's movements, as people with constant speed in the camera range will soon go outside. Multiple cameras can effectively solve this problem. Of course, the system software and hardware architecture needs some changes.

Again, the current system does not achieve the free of change of the final video, for the camera and testting conditions limit, the system currently only tested under limited conditions.

Finally, We believe that as the disciplines and principles of image processing technologies becoming more and more mature, less power consumption, faster detection technology, in the future, it will slove some problem such as camera image can not be real-time, and also the technology of human-computer interaction inputed by people moving will be applied to the real life and production more widely and quickly.

6. References

[1] R. C. Gonzalez and R. E. Woods, "Digital Image Processing," 2nd Edition, Prentice Hall, NJ, 2002.

[2] G. Agam, "Introduction to Programming with OpenCV," Vol. 2, 2007. http://www.cs.iit.edu/~agam/cs512/lectnotes/opencv-intro/index.html

[3] L. Schomaker, J. Nijtmans, A. Camurri, *et al.*, "A Taxonomy of Multimodal Interaction in the Human Information Processing System," Vol. 6, 1995. http://hwr.nici.kun.nl/~miami/taxonomy/taxonomy.html

[4] C. Gu and M. C. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 5, 1998, pp. 572-584.

[5] W. B. Liu and B. Z. Yuan, "From Actual Reality to Virtual Reality," *Chinese Journal of Electronics*, Vol. 2, 2001, pp. 100-105.

[6] J. X. Sun and D. B. Gu, "A Multiscale Edge Detection Algorithm Based on Wavelet Domain Vector Hidden Markov Tree Model," *Pattern Recognition*, Vol. 37, No. 7, 2004, pp. 1315-1324.

[7] H. S. Zhu, "Communication between the Application Process and Implementation of Technology," *Computer Applications and Software*, Vol. 21, No. 1, 2004, pp. 118-120.

[8] L. Gao, Y. L. Mo and B. W. Zhu, "Adaptive Threshold Segmentation of the Image Edge Detection Method," *Computer and Communications*, Vol. 25, No. 5, 2007, pp. 73-76.

[9] G. F. Yin, "Threshold Method Based on Image Segmentation," *Modern Electronic Technology*, Vol. 23, 2007, pp. 107-108.

[10] C. Hui, S. He and J. L. Hu, "Video Images of Moving Target Detection," *Computer Age*, Vol. 8, 2006, pp. 19-24.

[11] G. Ge, F. H. Fan and J. X. Peng, "Background Image in the Sequence Alignment and Motion Detection Algorithm. Data Acquisition And Processing," *Journal of Data Acquisition & Processing*, Vol. 12, No. 2, 1999, pp. 164-166.

[12] Geng for the East, L. M. Song, "Accessibility Camera-Based Interactive Techniques," *Computer Applications*, Vol. 27, No. 9, 2007, pp. 2087-2090.

[13] Z. Hong, Z. H. Wang and Banyan military. "Consistency of the Video Object Based on Motion Segmentation," Naval University of Engineering, Vol. 19, No. 4, 2007, pp. 91-93, 110.

[14] N. Yao, S. T. Li and J. X. Mao, "Computer Image Processing and Recognition Technology," Higher Education Press, 2005.

[15] J. L. Zhang and J. Liu, "Based on Wavelet and Morphology of Image Segmentation," Chinese People's Public Security University (Natural Science), Vol. 2, 2007, pp. 65-67.