The Research and Implementation of Interactive Video System Based on Android

Yang Li [1, a] and Ping Shi [2, b]

[1] Communication University of China, Beijing 100024, China

[2] Communication University of China, Beijing 100024, China

[a]332109990@qq.com, [b] shiping@cuc.edu.cn

Keywords: Interactive video system, C/S framework, mobile terminal, server, interactive video player and editor, KCF, Android.

Abstract: Interactive video system based on Android is a new mobile software application which enables the mobile phone users to make and browse interactive videos by themselves. This system can automatically display the corresponding things' additional information by users clicking them in the video and also offer the service that users can release and edit interactive videos by themselves. This paper will introduce the core technology of interactive video system and achieve tracking the targets' area which users are interested. Finally, a new interactive video system will be designed based on C/S framework. Combined with tracking algorithm, interactive video technology, and data communication methods, the interactive video player and editor will be designed in this system.

1. INTRODUCTION

With the continuous development of internet technology, mobile video application has become a hot spot. For users, it is not satisfied with the traditional way of watching video. Instead, users want more information from the video and want to interact with the video contents. For that reason, Interactive video arises at this historic moment. Interactive video is a kind of new multimedia that can automatically display the corresponding things' additional information when users interact with the video's target area [1]. Using interactive video technology will give ordinary video target more valuable information, which can be words, pictures, video, commodity connection and so on. As a result, this system can be used in many fields such as advertisement ,education ,and online entertainments [2].

Interactive video system based on Android mainly contains three types of information: user information, video information and target information. Target is the area which users are interested in. Its information includes location, time, text, images, and other video connections. This system can help users to make two different kinds of target area in the video. One is still area, the other is tracking area which can follow the moving target automatically.

Interactive video system based on Android contains two parts: mobile terminal and server. Mobile terminal realizes interactive video playing, interactive video editing, file transfer, user login, user Register and other functions. Server realizes the target tracking, interactive videos storage, file transfer, background management and other functions.

The rest of the paper is arranged as follows: Section 2 describes the system architecture. In Section 3 details of implementation methodology is discussed. Section 4 is about result and discussions. Conclusions are presented in Section 5.

2. SYSTEM ARCHITECTURE

Interactive video system based on Android is designed to share and edit interactive videos for mobile users. The system architecture is illustrated in Fig. 1. Mobile terminal is responsible for interacting with users. It has four modules: Interactive Video Player Module, Interactive Video Editor Module, User Management Module and Mobile Network Interface Module. Server is responsible for

receiving and processing data from mobile terminal. It also has four modules: Interactive Video Database Module, Video Tracking Module, Server Management Module and Server Network Interface Module.
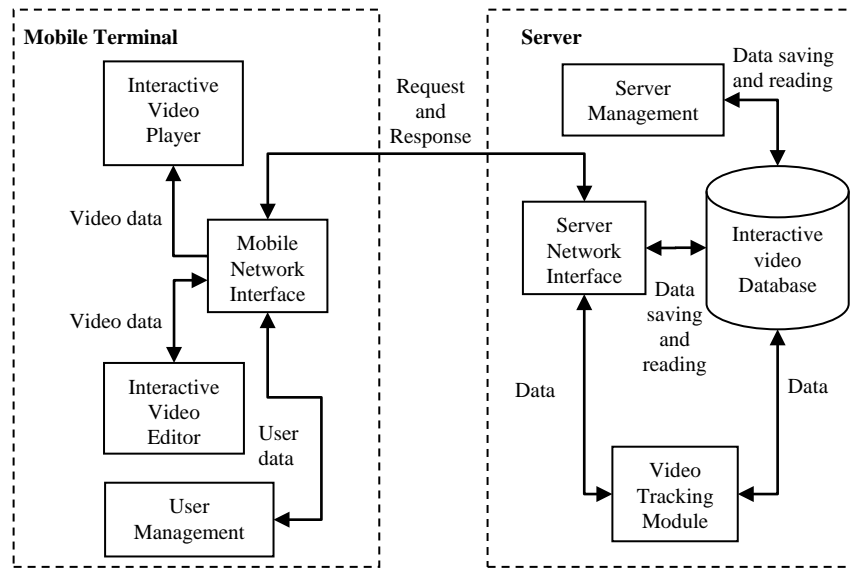

Fig. 1. Interactive video system architecture

2.1 Mobile Terminal

As Fig. 1 shows that mobile terminal is mainly used for user interaction and data transmission . As there are many types of android mobile phones and many mobile phones have low resource configuration, complex tasks should be completed by the server, Such as video tracking and data storage. Android tracker perform worse than the same type of tracker in PC [3], so mobile terminal should focus on the way to interact with users ,data transmission and UI design. Its four modules will be discussed in the following paper.

• Interactive Video Player Module is different from traditional video player. It not only plays the streaming video, but also provides the interactive function. After users start to watch a video, the player will get the video from server and get the additional information at the same time. Player has two overlapping views. One is to show the video at the bottom, the other is to draw the target area and judge whether the area is clicked or not which has transparent background on the top. If the target area was clicked, the player would be paused and there would be a dialog displayed with this target's additional information. The information can be words, images' addresses and videos' addresses . When the dialog is displayed, the images and videos will be downloaded and shown automatically. These processes are shown in Fig. 2.
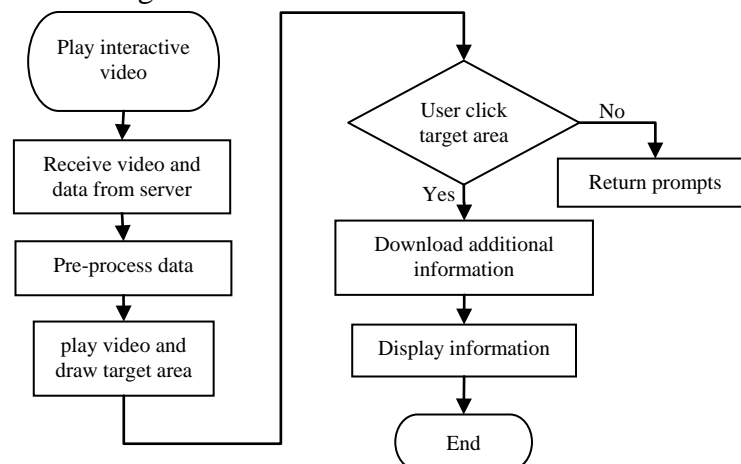

Fig. 2.   Activity diagram of interactive video player

• Interactive Video Editor Module has two functions. One is releasing videos, the other is editing targets' information online. This two functions are similar. Posting a video demands users to add the information and select the targets which can be still or tracked. After posting a video to the server, the server will analyze all the targets and execute tracking. All the tracking results and information including words, images and videos will be saved in the server database. When users start editing a video, the editor will download all the information, show the information and play the video online at the same time. The editor help user to modify all the targets' position and duration online if the tracking results are not accurate. It provides user a real-time tracking function. When user modify the target position in the video at some time point, the editor will send data to server, the server will execute tracking for this target in the appointed duration and give back results to the editor. Editor will analyze the results and show them to user. After finish editing all the contents in the video, the editor will send this new data to server for updating the database. These processes are shown in Fig. 3 and Fig. 4.
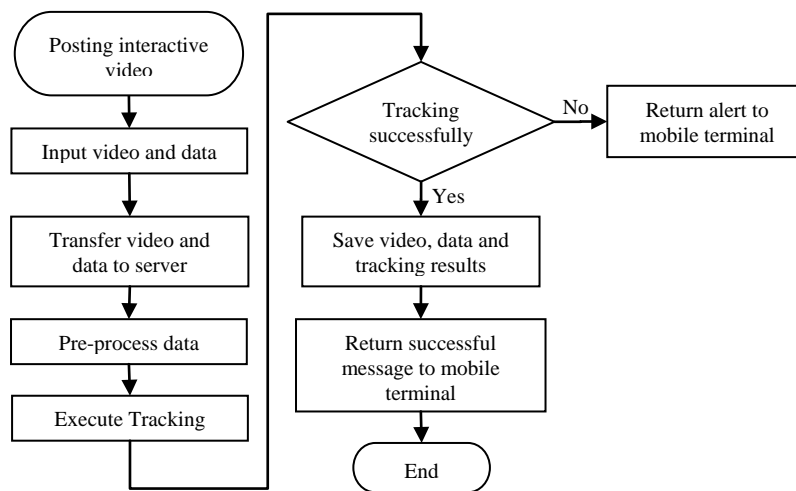
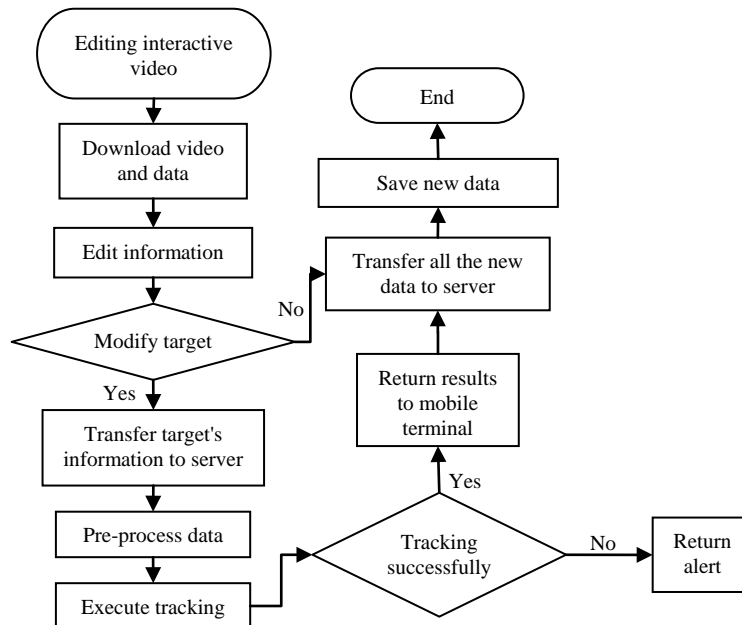Fig. 3. Activity diagram of posting interactive video

Fig. 4. Activity diagram of editing interactive video

• User Management Module is the entrance for the application. This module help users to register and login the application. When users login the application successfully, the server will make a signature automatically and send it to mobile terminal for a long-time login function.

• Mobile Network Interface Module is paired with Server Network Interface Module. The Interfaces provide data transmission function between each other. The network interfaces Use the

HTTP [4] transport protocol to transport data and use JSON(JavaScript Object Notation) to Encapsulate data [5]. JSON is a lightweight data-interchange format. With the characteristics of easy to read and writing quickly, it can exchange data between different platforms [5]. Mobile Network Interface Module mainly transfers the video and additional information to the server from users and waits for response to analyze.

2.2 Server

As Fig. 1 shows that interactive video server is mainly used for data storage ,target tracking and information management. The data from mobile terminal will be converted to general results from JSON format. Then the targets asked to be tracked will be sent to the tracking module. After finish tracking , all the information will be saved in the database. The server management has a web page to help administrators to manage all the information conveniently. Its four modules will be discussed in the following paper.

• Interactive Video Database Module is responsible for data storage and reading. Considering the types of information, the database at least needs five tables to save the data. The five tables is Video Table, Target Table, User Table, Admin Table and Comment Table. Video Table is used to save all the video data. Target Table is used to save the additional target information from videos. User Table is used to save users' information. Admin Table is used to save administrators' information. Comment Table is used to save the videos' comments from users.

• Video Tracking Module is used to achieve the goal of automatic tracking. After the tracking module got the target's starting position and time information, it would execute tracking algorithm, process the results and send them to database finally. This module is designed to get the tracking results quickly and doesn't need the results very accurate, because the user can use editor to modify the target position. So the tracking algorithm and tracking mode needs to fit the characteristics above. Moreover, if the video has a long time and the target area exists in a long duration, there will be many position results for the target, because a video has 24 frames per second generally. The interactive function doesn't need so many results to judge, so reducing the target's results is necessary.

• Server Management Module is designed to manage all the data in server. Administrator can use browser to communicate with the server to manage the user and video information directly.

• Server Network Interface Module is paired with Mobile Network Interface Module. They have similar functions. This interface receives data or requests from mobile terminal and executes corresponding operations in order to send data or responses back to mobile terminal.

3. DETAILS OF IMPLEMENTATION METHODOLOGY

Interactive video system based on Android was implemented by Android SDK Revision 22.6.2. The server used JAVA SSH (Struts + Spring + Hibernate) [6] as architectural structure. SSH is an integrated framework of Struts + Spring + Hibernate, which is a more popular kind of open source Web application framework currently [6]. Based on the above framework, the server will simplify tedious repetitive code in data access layer and also could be extended and maintained very easily. The database will use MySQL database as the platform which is a relational database and Very suitable for the data in this system. The system also used OpenCV [7] to support the tracking algorithm and used JNI [8] technology to link server to tracking module. The full name of OpenCV is Open Source Computer Vision Library which Consists of a series of C function and a small amount of C++ class [7]. It realizes the image processing and computer vision for many common algorithms. JNI is an acronym for Java Native Interface. It provides several API for Java and other languages communication (mainly for C&C++) [8].

3.1 Server and database management

Interactive video system server used JAVA SSH (Struts + Spring + Hibernate) architecture, the user's browser can communicate directly with the server J2EE (JAVA 2 Platform, Enterprise Edition) [6] communication components by Server Management Module. SSH divides system into four layers: the presentation layer, business logic layer, data persistence layer and domain module layer [6]. Struts is responsible for the MVC (view, controller and model) separation and take control of the program logic of jump. Hibernate is responsible for the efficient implementation of data access in persistence layer. Spring is mainly responsible for management and scheduling of Struts and Hibernate [6]. In the process of actual system setting up, we implement the java object according to the needs of the business design transaction prototype, then we write Hibernate DAO (Data Access Objects) to realize the binding of Data Access classes and database. In the end we design Spring management and scheduling rules to manage the struts and Hibernate.

Considering the interactive video and the corresponding relationship between the user and the target object, we use MySQL database as the database management which is a relational database management system. There are five main tables in the interactive video database now which is Video Table, Target Table, User Table, Admin Table and Comment Table.

3.2 Data transmission way and network interface

Mobile applications has two main ways to realize the communication with the server in roughly. One is HTTP communication mode and the other is Socket communication mode. The communication mode of HTTP is using HTTP protocol(Hypertext Transfer Protocol) to communicate. Its most remarkable characteristic is that after the client sends each request to require server echo response, the server will take the initiative to release the connection at the end of the requests. Considering the function of this system, the mobile terminal doesn't need to keep connection with server in a long time, so we choose HTTP communication protocol to save the server resources.

JSON (JavaScript Object Notation) is a lightweight data-interchange format. With the characteristics of easy to read and writing quickly, it can exchange data between different platforms [5]. As a result of the targets of the video are numerous, the system should choose simple description and easily parsed transmission format. So JSON is used to encapsulate data from users.

The two network interfaces communicate with each other and transfer images or videos through HTTP communication protocol. At the same time the additional data will be encapsulated into JSON format and sent in a asynchronous thread. The details of interfaces' functions are as follows: USER_LOGIN, USER_REGISTER, GET_ALL_VIDEOS, PUBLISH_VIDEO, PLAY_VIDEO, EDIT_VIDEO, DELETE_VIDEO, ADD_TARGET, EDIT_TARGET, DELETE_TARGET, PUBLISH_COMMENT, GET_ COMMENT, UPLOAD, TRACKING.

3.3 Tracking module

Video tracking module was implemented with C++ and JAVA language. The tracking algorithm will be compiled into DLL(Dynamic Link Library) and the server use JNI to link server with DLL. DLL is short for Dynamic Link Library, which could package different functions for different system to use. The tracking algorithm also needs OpenCV to support the basic function. On the whole, the server use JNI to connect the tracking algorithm DLL. The DLL execute tracking algorithm and send back results to server.

Video tracking module design involves two parts: the tracking mode and the tracking algorithm. As the above paper said, this module should be suitable for two demands. The first one is tracking quickly. The second one is that all results should get reduced for saving conveniently which don't need to be very accurate. Considering this situation, the system uses a special tracking mode: interval frames to track and saves only one position for one second. The system use interval frames to track for getting high speed. According to different video quality, the interval number can be adjusted. The interval number can be zero which means tracking frames one by one. In this system we use three for

interval number. With the interval number and the video frame rate, we could calculate the number of results for one second. We choose the first frame's result as the final one for this second generally. For example, the interval number is 3, the video frame rate is 24, so the number of results will be 8 for one second. We choose the first result as the final one for this second.

For fast tracking the video target, the system use KCF(Kernel Correlation Filter) [9] tracking algorithm, which is proposed by João F. Henriques, Rui Caseiro, Pedro Martins, Jorge Batista. KCF is an identification approach for tracking. This kind of method is usually training a target detector in the process of tracking. Using the target detector to detect the predicted target in the next frame and then using the new test result to update the training set in order to update the target detector. Different with other tracking algorithm , KCF Used ridge regression to train target detector and convert matrix calculations to vector Hadamad product successfully based on the nature that cycle matrix can be diagonalizable in Fourier space which greatly reduces the computational complexity, improves the operation speed and meets the real-time requirements [9]. It also presents a way that multi-channel data can be incorporated into the algorithm. KCF uses HOG [10] as the multi-channel data. Compared with TLD [11], Struck [12], CT [13] and other tracking algorithms , KCF can already reach competitive performance by operating on raw pixels alone. Replacing pixels with HOG features allows the KCF to surpass even TLD and Struck by a relatively large margin. KCF make the tracker run at hundreds of FPS and can be implemented with only a few lines of code [9]. With good precision and high speed, the system use KCF as the tracking algorithm.

When the target's data is sent to tracking module, the module will use JNI to link to the DLL which has the tracking algorithm. JNI will convert the data format and send it to DLL. DLL will pre-process the video with the data and track video by Interval selected frames. After finishing the tracking, results will be reduced by the tracking mode above. Then JNI will get the results back to server and output them to database or mobile terminal. These processes are shown in Fig. 5.
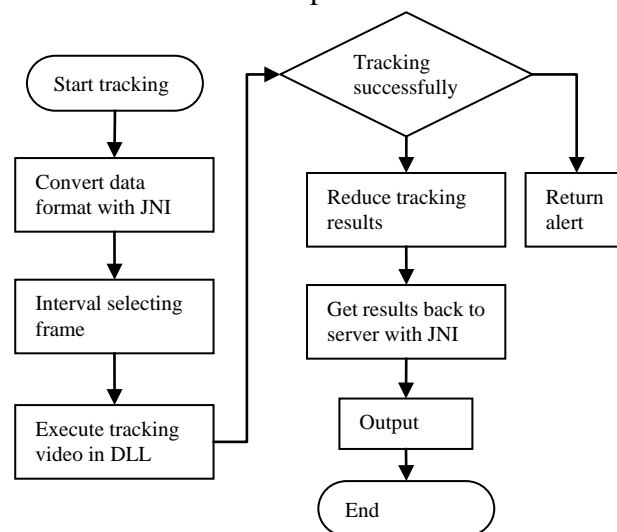

Fig. 5. Activity diagram of tracking module

3.4 Interactive function

The interactive function is mainly concentrated in the play and edit module. The interactive way is by users clicking the target area. If users are not interested in this target, this target's additional information won't be shown to users. Both play and edit modules have a interactive video player for users watching. The player is a streaming media player with using HTTP protocol which could play videos by a streaming format online [4] [14]. The player has two layers: presentation layer and data processing layer. Once the player starts to play a video, there will be two threads got opened. One is receiving the video file from server, the other is getting all the video additional information and processing the data. From the above paper we know that the tracking target's positions are saved by second, namely, the positions get reduced and there is just one position in a second. Because a general

video at least has 24 frames for one second, we need to expand these positions before showing these target areas.

In data processing layer, the system will divide one second into many periods and do the interpolation processing between each two positions. Then the system can get every position in that period .For example, one second divided into 40 periods. It is 25ms for every period. Then we insert 39 points into two positions and do the interpolation processing. After that the 40 periods will have their own position and we will get 40 new positions in one second. By this way the system can expand the target's position and make the area shown smoothly.

In presentation layer, the player has two overlapping views. One is to show the video at the bottom, the other is to draw the target area and judge whether the area is clicked or not which has transparent background on the top. After finishing processing the data, the bottom view will start a new thread and play the video. At the same time, the transparent view on the top will draw all the targets area with the processed positions in the appointed period. When user clicks the target area, the view will get the position that user clicked and compare it with all the targets' positions at this current time point. If the position users clicked belong to some target's position, the information will be displayed to users in a dialog.

4. RESULT AND DISCUSSIONS

In this paper, an interactive video system based on Android has been developed. Users could use android mobile phone to install this application. The application showed the list view of videos on the main page(Fig. 6a). User could release and edit video or target information in the editor module(Fig. 6b). User can change their username, password and head photo in the user center(Fig. 6c).

By using this interactive video application, a user could make their own interactive videos and put interesting words, images and videos on the target area, which can make others know your video better. The application offer a platform for people to share interactive videos and could change the single way of people watching videos. The advertising companies and Network video companies could use this system to create much more commercial values.

The application provides the search function to help people to make friends by telephone number. It also provides user make comments under the video they watched. When a user prepares to post an interactive video, he doesn't need to wait for the tracking results. All needs to be done is just to select the target, choose time and post it. The server will finish the tracking on the background automatically. Other users will see your interactive video directly with tracking effects.
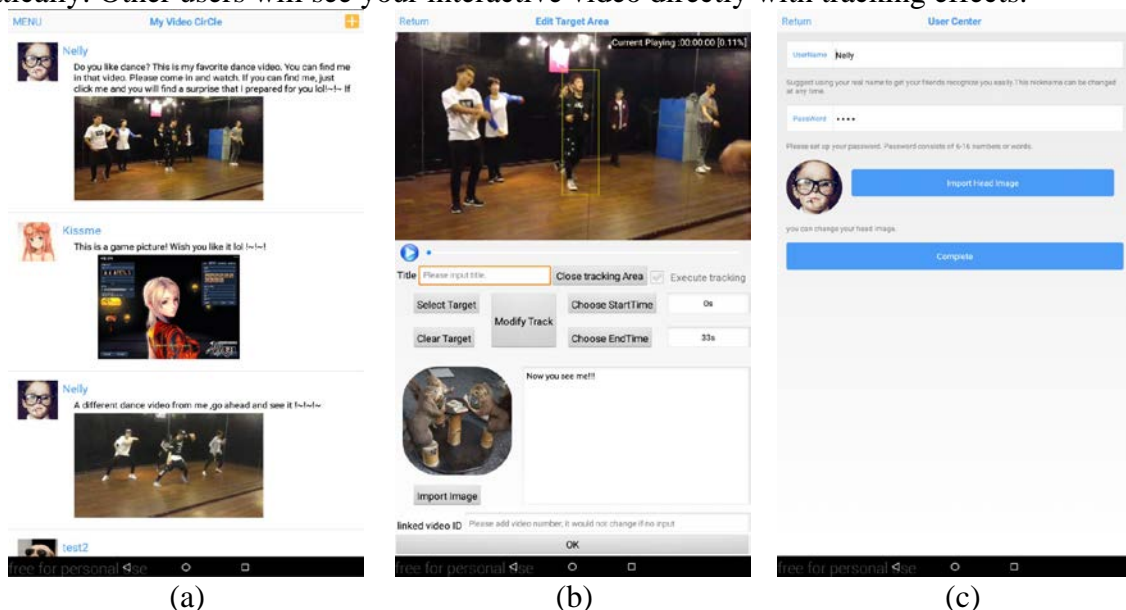

(a)   (b)   (c)

Fig. 6. The view of interactive video application

5. Conclusions

Based on the design, implementation, and testing of the system, it that can be concluded that :

1. An interactive video system based on Android has been completely developed. The system contains mobile terminal and server. The system was built by using SSH framework, OpenCV, DLL and JNI.

2.Based on the testing results, the system offered a platform for users to make, share, save and browse their own interactive videos.

3.We proposed a new solution for interactive video system, including the interactive mode, tracking implement, data transmission and Interactive video's display mechanism. The system realized the interactive video player and editor. The tracking mode and algorithm still have some shortcomings. The high speed movement or non-rigid deformations will affect the tracking precision. The improvements of tracking mode and algorithm are necessary, which will make the system more convenient to mobile users.

Acknowledgments

References

[1] N. Sawhney, D. Balcom, and I. Smith, Authoring and Navigating Video in Space and Time, In IEEE Multimedia, 1997, 4(4):30-39

[2] G.H.Cha.Object-Based Interactive Video Access for Consumer-Driven Advertising[J]. Proceedings of the 8th International Conference on E-commerce and Web Technologies, 2007, 222-228

[3] Wang Jue, Yuanming Wu. Multi-target real-time tracking technology based on Android platform. University of Electronic Science and Technology of China. 2016. 05

[4] Lizhi Li, Leyu Wang, Yilun Huang. Streaming media player system software design based on HTTP protocol. Zhejiang university. 2006. 05

[5] wikipedia.org.JSON[EB/OL].http://en.wikipedia.org/wiki/JSON,2010-2-5

[6] Tao Yizheng, Wu Zhijie, Tang Dingyong, et al. Research of application framework technology based on J2EE [J]. Computer Engineering and Design, 2007. 28 (2):826-828

[7] R.Laganière. OpenCV 2 Computer Vision Application Programming Cookbook: Over 50 recipes to master this library of programming functions for real-time computer vision[M]. Packt Publishing Ltd, 2011,1-13

[8] S. Lee, J. Jeon. Evaluating performance of Android platform using native C for embedded systems[C]. 2010 International Conference on Control Automation and Systems, 2010: 1160-1163

[9] João F. Henriques, Rui Caseiro, Pedro Martins, Jorge Batista "Kernelized Correlation Filters "ECCV 2012, TPAMI 2015

[10]HOG: one   of   target   detection   of   image   features   extraction. http://blog.csdn.net/liulina603/article/details/8291093

[11]Z. Kalal, K. Mikolajczyk, J. Matas. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012,34(7): 1409-1422

[12]S. Hare, A. Saffari, P. Torr. Struck: Structured output tracking with kernels[C]. 2011 IEEE International Conference on Computer Vision (ICCV), 2011: 263-270

[13]Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang" Real-time Compressive Tracking" ECCV 2012

[14]Can Yang, Yongyan Li, Jionglong Chen. A New Mobile Streaming System Base-On Http Live Streaming Protocol. Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on. 2011. 10