ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

AUTOMATED SYSTEM FOR CONTROL OF MULTIPLE CAMERAS WITH RESPECT TO THE ACTIVE SPEAKER

Martin Sivý, Miroslav Michalko
Department of Computers and Informatics, Technical University of Košice, Slovak Republic

DOI: 10.5281/zenodo.884126

ABSTRACT

In this paper, the research on automated system for control of multiple cameras is presented. The principal goal is to enable switching of cameras based on the position of active speaker. For the purpose of this system it was crucial to design PTZ cameras positioned in specific location, i.e. local government hall. In total, three independent system components are created and communicate with one another using REST calls. Each of these components deals with a specific logic part of the whole system. The first component recognizes the active speaker by monitoring the input along using OCR system and consequently outputs name of the speaker. The second component is for the management of members stored in the system and accepting recordings of active users. The last component is the most user oriented. It has a graphic interface in which the user can define cameras position beforehand and manage members of the system. The most critical function is the automatic behavior of cameras.

KEYWORDS: camera, multi-camera system, OCR, person recognition, PTZ.

I. INTRODUCTION

Every day hundreds of events from around the world are recorded to increase the security (objects, buildings, people), awareness (news, weather, discussions) or entertainment (concerts, TV shows). In current modern world, people have a lot of opportunities to watch many of these events through the Internet, almost without any delay. Each of us can decide how to spend the time, whether its sports events, music, education or something else. There are plenty of choices. Watching video stream is one of these. Of course, behind the video stream, there is a multi-person team of people. These people make possible that the viewer gets to the essentials and nothing important is missed. They are directors, cameramen, sound engineers etc.

The goal of this paper is to simplify the work of local television team consisting of cameramen and technical members of the city council. This is to be done using a fully automated PTZ camera systems, while shooting is to be carried out with respect to the active speaker. As an active speaker we consider person who speaks and all the attention is centered on him or her. There are many existing solutions how to control PTZ camera systems using software or hardware keyboards. None of these solutions offer the opportunity to determine the exact active speaker and to direct the camera to this person. Many of such software solutions are unnecessarily complicated to control. This research should facilitate the work of the technical members of the television staff, since they will no longer be required to carry all the equipment.

This very research is one of three separate researches designed to improve and simplify the work of local television companies, as they will have less workload in securing the broadcasting and filming of city council gatherings. The overall system solution, being the unification of these three works, is to provide the complete automatic support from shooting city sessions to live observation by city residents. Citizens will be able to watch live streaming from the home, using mobile applications supporting the two most popular mobile platforms (iOS, Android). Further, commenting and voting is to be implemented. Specifically, this work is intended to ensure fully-automatic live video transmission with respect to the active speaker. All three researches are interconnected.

II. ANALYTICAL CONSIDERATIONS

The topic of this research was divided to several smaller parts. We will present the analysis of PTZ camera systems. PTZ camera systems are actually cameras that can be positioned horizontally, vertically and they can be controlled remotely by using hardware or software keyboards. In addition, we present the protocols used to

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

communicate with PTZ camera systems, real-time broadcasting systems and also an overview of similar existing solutions that are currently used in city councils. There is a large number of companies in the world dealing with hardware development as well as software for PTZ camera systems and video transmission through the Internet.

Control of PTZ camera systems

Positioning the PTZ camera systems have the ability to rotate and zoom in different positions, making operation much easier. Control of PTZ camera systems can be summarized to these main groups:

- control of PTZ camera systems using software (software keyboard),
- control PTZ camera systems using hardware (hardware keyboard).

Communication paths between peripheral devices (camera systems) and controllers (computers) can be realized using metallic cables, optical cables or wireless technology such as radio signal.

Hardware keyboard

This type of PTZ camera control is very expensive. The market offers a large range from primitive controllers to professional desks that can easily handle multiple PTZ cameras at the same time and may even be equipped with a screen where user can directly see the output image from the camera.

Software keyboard

The second type used to control PTZ camera systems is the software keyboard, which can by functionality fully replace the hardware keyboards. These kinds of keyboards fulfil the same task and it is easier to acquire them. Price is many times lower when compared to hardware keyboards. As a part of this study it will also be a goal to design and implement this type of keyboard.

Communication protocols to control PTZ camera

Of course there must be a mechanism for controlling PTZ systems. Several companies have dealt with this problem. Then we will describe some of the most used solutions, specifically protocols:

- Ernitec,
- VISCA,
- PELCO-D.

Ernitec protocol

This protocol serves to control the positioning camera systems. It is a serial asynchronous one-way simple protocol. Its basic features are:

- possibility of control up to 32 PTZ camera systems (Ernitec ICU / BDR510 / 550),
- horizontal/vertical direction, zooming in/out, focusing,
- ability to set up and call up to 99 presets for one camera system,
- ability to adjust the camera's rotation speed.

VISCA protocol

The VISCA protocol [1] is one of the professional protocols for controlling PTZ camera systems designed and implemented by SONY. Protocol is based on RS232 [2] serial communication at 9600 bit/s. When the party sending commands such as a computer is called a controller and the receiving party, for example the EVI D70 camera system calls a peripheral device. In this protocol, up to 7 peripheral devices, such as this camera system (EVI-D70), can be connected to one controller via the RS-232C / RS-422 standard [3]. The RS-232C / RS-422 parameters are as follows:

- communication speed - 9600 bps / 38400 bps,
- data byte - 8,
- initial byte - 1,
- end byte - 1,
- parity - none.

The basic unit of VISCA communication is the so called packet. The first byte of package is called header and includes the sender and recipient address. To use broadcast, this byte is set to a hexadecimal value of 88H and processed by all peripherally connected devices. If there is a requirement to execute an instruction from only one camera system, such as one having assigned serial number 4 on the controller side (control unit), this hexadecimal value is set to 84H.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

PELCO-D

This protocol is also a well-known and advanced protocol in the CCTV industry. The protocol has one start byte, eight data bytes, and one ending byte. The normal transmission rate is 4800 baud per second [4]. In order for the camera system to function properly, it is necessary to send messages in the correct format, the structure of the PELCO-D message log is:

- synchronized byte,
- camera address,
- command 1,
- command 2,
- command 3,
- data 1,
- data 2,
- checksum.

Comprehensive real-time video streaming solutions

There are, of course, solutions such as transmitting video directly from cameras to a server from where it will be further disseminated to the viewer. There are two solutions, LiveShell.X and the more expensive Teradek Sphere HDMI. Both solutions are used in the professional sector.

LiveShell.X

LiveShell is a high-quality product from the Japanese production designed for live video streaming. HDMI cable is used on the input side to connect to the camera and on other to Internet. Once properly configured, this device automatically logs on to the server and starts transmitting the recorded video signal to the server in real time. This very type, LiveShell.X is the latest product of Cerevo and also belongs to the top-of-the-line models of this company. It allows real-time video streaming without the need to connect the computer, making video streaming even easier.

Teradek Sphere HDMI

This device offers 360° real-time video streaming without the need of a computer device. It mainly utilized iOS support and Apple iPad Pro, where it installs a mobile application that can control this device and thus transmit live video. It allows users to broadcast 360 ° video in high quality MP4 format directly on popular streaming platforms such as YouTube, Wowza or RTMP server. The application itself can optimize the image so that the quality of transitions is as good as possible and the output is of the highest quality.

Systems utilized by city councils

The next part deals with the analysis of the control systems used by city councils. Two solutions are mentioned, namely H.E.R. System and MINISTER. In Slovakia there is a greater occurrence of the H.E.R. This system is also being used by the city of Poprad. The MINISTR system is formerly used in the Czech Republic.

H.E.R. system

H.E.R. is a voting system designed for both local and city councils. The system is variable and can satisfy customers in different industries and is ready to meet all their requirements. The company is able to customize the customer with a small deviation and "tailor" a unique solution for the customer. This solution is available in two major variants, either fixed or mobile, taking into account the customer's space options. The system is simple to operate and equipped with voting units, smart card readers and controls. If customers have special requirements and need to have something tailored, the company is open to new ideas and does not have issues of "cutting off" the exact solution for the individual customers. This keeps the system up and adds the necessary new functionality [5].

MINISTR

The MINISTRY voting system is primarily intended for city councils and available for a better price than other competitive solutions. The system automatically archives the entire session, including audio recording. It is and is mainly developed for the needs of the Czech market [6]. It has functionality for simplified search of individual records. Versions of this solution:

- Ministr ® 3,
- Ministr ® 4,

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

- Ministr ® 5,
- Ministr ® 6 Wi-Fi.

III. ON THE SOLUTION

The overall solution of this research is made up of three separate parts that communicate together with their own REST API. The whole project was adjusted to meet the requirements of the city council in Poprad. These three main parts, namely:

- PTZ camera control,
- identification of active speaker,
- server part.

The Java programming language was selected and also an objective language that is easy to read and edit for future use and extension. When designing these programs, a great deal has been taken of the security of the whole system. Also, its implementation may be multi-threaded. In this research, it is utilized to allow operation in multi-threaded mode just in the automated control of CCTV systems.

Server part

The large funds are spent on securing data and preventing unauthorized access to information systems. This application will handle sensitive data, and its exploitation would, at the least, disable the effective upload of a record from a city council, it is necessary to secure this application and authorize users. The application itself will run on a distributed Heroku system. Heroku itself uses encrypted communication with HTTPS, so readability of third-party data sent and received is almost impossible. The PostreSQL database is also secured by the name and password. In order for the user to use the features of proposed application, it is necessary to prove the correct combination of username and password. Passwords are, of course, not stored in a readable form but are protected in case database would be stolen. For this case, passwords are hashed and Salt is applied (see Fig. 1).

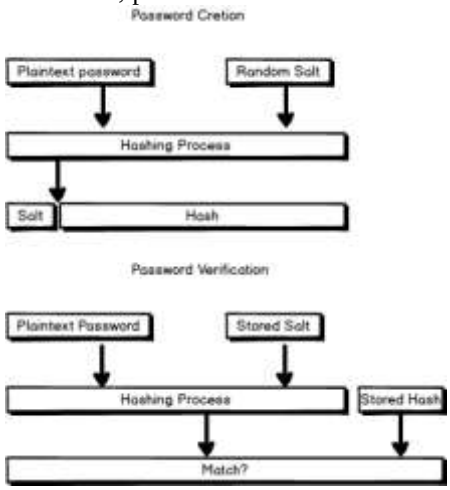

Figure 1. Password creation and verification with use of Salt

Salt is used precisely because even if the whole database was stolen and the attacker would try to use the hash table, it would be useless because he would not get complete records. Salt adds characters such as capitals, numbers and special characters to the password. The password is then encrypted. It should also be remembered that, when verifying the password entered by the user, Salt should be removed to be validated. An attacker would have to deal with each stored password separately, which would take up a number of times to solve all the passwords.

Therefore, the user must be logged in to allow the user to upload the data. Once a token has been successfully obtained, this token is valid for next 24 hours. After this time, if the users want to use the system again, they must do login process again. Token is a substantially long, indivisible character string that is unique throughout the user's system, and can have encrypted and added data such as the user name and the application for which it is created. This 24-hour period is definitely longer than the duration of the ordinary city council meeting and therefore it should not be in any way annoying for the user. For each additional request to the server, the user

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

sends this token so that the system knows users identity. In the case of failure of authentication, the user is informed about this fact in the form of error message and error code.

PTZ camera control

Software solution, as part of the whole project is the only one that contains a graphical interface. Its use is aimed to the user, a dedicated city technician, who takes care of the smooth running of the city council gathering. In the design of solution, it was considered that the editing will only be used if the members of the city council change. Normally, the user only switches on the so-called auto mode and the system will handle all connected PTZ camera systems themselves. In the first step, it was necessary to create a user login window. This window has two text fields for entering a name, password, and one sign-in button. The login name is entered legibly but for a higher security level, the password entered by the characters will only appear as dots. If the login was unsuccessful, an error message is displayed.

After successful login, the program switches to so called "Main user screen" (see Fig. 2).

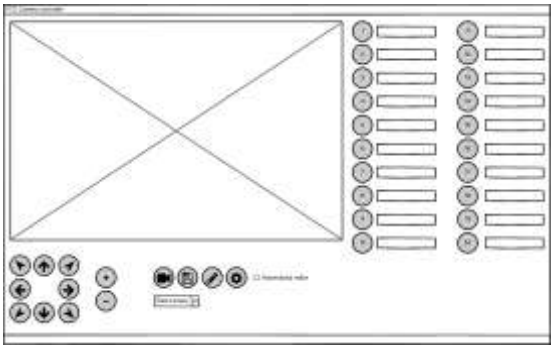

Figure 2. Main user screen PTZ

This main screen can be further divided into three larger logical units:

- display of the received video camera signal,
- PTZ camera systems control and automatic mode trigger,

display the current list of city council members with their seats.

Control signals

It is essential to seek the possibilities of enabling the communication between the program and the connected camera systems and thus enable systematic control of these systems. The most supported protocol for control is the VISCA protocol. This VISCA protocol is supported by SONY EVI-D70 camera systems, which will be used to test the functionality. This protocol allows to involve four camera systems needed for a solution in the premises of city council gathering. If the camera system receives an incorrect message, it does not perform any action. In case of VISCA protocol it must send these messages in the form of a byte. Hexadecimal access to instructions were selected and can be read from the Table 1. In Table 1, the individual VISCA instructions are described for camera system movement and positioning. The proposal was based on the original documentation on the VISCA protocol.

Table 1. Individual instructions for VISCA protocol

Instruction name	Hexadecimal VISCA instruction code
right	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x02, 0x03, 0xff}
left	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x01, 0x03, 0xff}
down	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x03, 0x02, 0xff}
up	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x03, 0x01, 0xff}
up left	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x01, 0x01, 0xff}
down left	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x01, 0x02, 0xff}
up right	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x02, 0x01, 0xff}
down right	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x02, 0x02, 0xff}
stop instruction	{0x81, 0x01, 0x06, 0x01, 0x14, 0x14, 0x03, 0x03, 0xff};
zoom in	{0x81, 0x01, 0x04, 0x07, 0x02, 0xff}
zoom out	{0x81, 0x01, 0x04, 0x07, 0x03, 0xff}
zoom stop	{0x81, 0x01, 0x04, 0x07, 0x00, 0xff}
reset	{0x81, 0x01, 0x04, 0x3f, 0x00, 0x01, 0xff}

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

set	{0x81, 0x01, 0x04, 0x3f, 0x01, 0x01, 0xff}
recall	{0x81, 0x01, 0x04, 0x3f, 0x02, 0x01, 0xff}

These instruction sets will be called when user presses the appropriate software buttons to control PTZ camera systems. Thus, when pushing these buttons, the corresponding direction of rotation or the setting or eventual invocation of the preset position of the camera system are sent. VISCA protocol uses only one instruction to set the movement of camera to the right, and the camera is still rotating at the specified speed to the right, up to its limit. Hence, it is necessary to stop the rotation at the desired moment. To stop running any of the instructions, special “stop instruction” is used. For simplicity, the associated instructions will be called once the button is pressed, and when it is released, the original execution of the instruction is interrupted by "stop instruction". This will allow the user to control the motion of the camera system smoothly and not just in any of the initially defined moving steps. If the user holds the movement button, the camera slowly turns until the user releases the button or the camera does hit its limit values. Cameras are connected from the computer in series. Each camera system has its own serial number within connected computer. As a rule, CCTV systems are numbered according to the order in which they are connected from computer. Of course, in order to send an instruction to a camera system, the program must know which port to send the instruction to. All the parameters for this transfer will be defined in the application and the user is only expected to select the correct serial port from the list of currently connected ports. The camera system is connected through the USB cable on the computer side and the ports are marked COMXX, where XX denotes a decimal number. RS232 standard will be used for camera system SONY EVI-D70CP.

Identification of the active speaker

The last part of the research includes proposal of a software solution for detection and recognition of the active speaker. Image of speaker is uploaded to the server part using REST call, from where it can be read by other applications. The current city council management system includes a screen which will be used as a clue. The procedure is as follows, the user, as the authority, logs in to this system to be authorized to record an active speaker. Login in is to prevent the abuse of the system and to prevent degrading the session record. Subsequently, this program automatically enters an endless cycle until it is shut down. The system will select the area where the active speaker is likely to be in proportion. OCR software recognizes name of the active speaker and uploads it using REST to the server part.

The pilot part for recognition of active speaker for this work is the use of OCR, JavaCV. It was necessary to find a solution other than acquiring the active speaker, since it is not possible to gain access to the system that is implemented and used by the council gatherings. System does not offer any public accessibility nor further integrations. This existing system (H.E.R.) uses the data projector for the information output. Data projector provides the following items:

- screen type,
- point number,
- meeting point,
- list of signed members,
- number of signed members,
- number of inserted cards,
- time of discussion of the current speaker.

In H.E.R. the following three main screen types are used where an active speaker can be found:

- submit,
- resolution,
- discussion.

The position of active speaker is still on the same place on all three screen types and therefore it is possible to proportionally crop the image where the speaker name is located. Even when the screen resolution changes, there is no problem with identification of an active speaker due to a fixed pixel-bound position. This section will help to avoid unnecessary use of OCR on the entire image, which would cost more computer time and get more unnecessary information. Thus, the likelihood of successfully recognizing the name of the active speaker is even higher. The text area is also selected with the words "submit:", "resolution:", "discussion:", and the name of the speaker that is behind the colon is sent to the server. The actual visual output from the meeting of the city council in Poprad is depicted in Fig. 3.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Figure 3. Discussion – output from H.E.R. system

IV. CONCLUSION

Presented paper aimed to ensure seamless access to the live broadcasting from city council gatherings. Thus, facilitate the workload of technical support and allow management of camera systems conveniently using proposed solutions. The basis for the successful design of the solution was acquired from the analysis of the control of camera systems using software or hardware keyboards and their possibilities of providing live broadcasting. Live streaming can be achieved using devices such as LiveShell.X or Teradek Sphere HDMI. There are many solutions on the market that deal with camera system management, yet none of these have been able to independently identify an active speaker.

Created solution proved to be able to control PTZ camera systems in any direction and use zoom, calling and storing the preset positions of camera systems. Also, another goal of this paper was met, i.e. identification of an active speaker, which was implemented by a separate application that must be launched on the same device as H.E.R. system. Identification of the active speaker was done by subtraction from the image output and using the possibilities of OCR library. The third, and last part, is the server part of this paper that manages user actions in the system and ensure data integrity. It runs on OAuth 2.0, ensuring a higher security standard [7]. Thus, the system created as a set of three independent solutions offers all the necessary functionality for controlling PTZ camera systems with respect to the active speaker, access to these functions are only available after valid login to prevent unauthorized manipulation with data. All the data regarding camera systems and city council members are safely stored in the PostreSQL database in the Heroku distributed system.

V. ACKNOWLEDGEMENTS

Paper is the result of the Project implementation: Competency Centre for Knowledge technologies applied in Innovation of Production Systems in Industry and Services, ITMS: 26220220155, supported by the Research & Development Operational Programme funded by the ERDF.

VI. REFERENCES

[1] MODULE, CAMERA. Technical Manual. Firmware Reference Manual, CSC, 2006.
[2] DE ALMEIDA PEREIRA ZUQUIM, L. D., et al. An embedded converter from RS232 to Universal Serial Bus. In: Integrated Circuits and Systems Design, 2001, 14th Symposium on. IEEE, 2001. pp. 91-96.
[3] CISCO: Cisco TelePresence PrecisionHD Camera, 2016, http://www.cisco.com/c/dam/en/us/td/docs/telepresence/endpoint/camera/precisionhd/user_guide/precisionhd_1080p-720p_camera_user_guide.pdf

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

[4] FRANZEL L.: What´s The Difference Between Bit Rate And Baud Rate? 2016, http://electronicdesign.com/communications/what-s-difference-between-bit-rate-and-baud-rate
[5] Hersystem: H.E.R. Systém, 2016, http://www.hersystem.eu/sk/her-system
[6] MINISTR, 2016, http://www.ministr.cz/hlasys3.html
[7] Ryan Boyd: Getting Started wirh OAuth 2.0. O'Reilly Media, Inc., 2012. pp. 80 ISBN 9781449311605.

CITE AN ARTICLE

Sivý, M., & Michalko, M. (2017). AUTOMATED SYSTEM FOR CONTROL OF MULTIPLE CAMERAS WITH RESPECT TO THE ACTIVE SPEAKER. *INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY*, *6*(9), 7-14. Retrieved September 4, 2017.