A Hybrid Real-time Video Streaming Remote Surveillance System Based on Mobile Robot

Zhi-jie XU[1] and Geng WANG[2]

[1]School of Software, Shanghai Jiao Tong University,Shanghai200240, China

[2]School of Software, Shanghai Jiao Tong University, Shanghai200240, China

[1]moonxiphoid@stju.edu.cn

Keywords: Video streaming system, Wireless mobile robot, Panoramic video.

Abstract. Recently, robot is more often introduced to warehouse cargo monitoring and surveillance thanks to rapid industrial automation. In the process, real-time video streaming system plays an important role. Improved accuracy of monitoring causes higher requirements for real-time, stable and efficient video streaming system, but mobile robot uses limited wireless transmission bandwidth in a warehouse, which is unable to ensure clear and real-time streaming simultaneously. This paper proposes a multichannel video stream transmission system to split multiple video data and distribute according to different transmission strategies, which attaches equal importance to clear and stable surveillance 360-degree panoramic video and real-time interactive video stream with low amount of data. By testing its performance with various parameters, it is revealed that the requirements for indoor monitoring of mobile robot are mostly satisfied by the system as it is able to transmit the panoramic audio and video streams at the code ratio up to 4mbps stably through common wireless network, as well as 1mbps interactive audio and video streams at the transmission delay of less than 800ms.

Introduction

Expanding industrial automation introduces more and more robots into warehouse to make cargo surveillance more intelligent and automated. For instance, Amazon.com, Inc. had 15,000 robots named Kiva running in its U.S. warehouse center. As a more efficient and labor releasing way than artificial surveillance, surveillance robot can offer scene replay and remote check, achieve 24h cargo surveillance uninterruptedly, and report any abnormality in a real-time manner, thanks to its continuous cargo surveillance and storage of information on cargos throughout warehouse.

For remote surveillance, video streaming system is essential. Remote video surveillance relies normally on a combination of fixed cameras and armed-with-camera robots, but this paper focuses on the latter. A robot can convert the information on warehoused cargos into video streams by means of camera to transmit the audio and video messages to the server at surveillance center in a wired or wireless way, where they are used by system or staff for making judgments. In the past, robot for automatic surveillance captured and acquired the scene of warehouse with multiple cameras in different directions, so it had to search video streams from multiple channels simultaneously to fetch the video for a specific moment without swift video change, making it troublesome for supervisors.

Recently, 360-degree panoramic video technology is developed into a more suitable way for information storage of mobile robot in charge of surveillance. Robot stores the panoramic videos of warehouse, so supervisor may check the cargos in a real-time and all-directional manner via PC or smart phone, which reduces the blind zones and targets out of view effectively. However, panoramic videos attribute their clarity to large network bandwidth, which is often difficult to maintain for wireless communication of robot in a cargo warehouse featuring complicated internal construction, and electromagnetic interference of cargos containing metals. The video pause, delay or frame loss of real-time surveillance video streams may be caused, especially under the real-time interaction for remote video connection. For instance, remote motion control of robot is not satisfied by video streams of large bandwidth and latency. This paper proposes a hybrid video stream transmission solution in which surveillance robot stores and transmits remote surveillance video streams in the form of 360-degree panoramic video, and extracts some videos for key zones to create a real-time video stream channel feature low latency and code ratio, in order to realize real-time interaction.

This paper will introduce the hybrid video stream solution and where it is suitable for in the second section, and present the algorithm and techniques used in the system for video processing and transmission in the third section. In the fourth section, we will give a brief analysis on the data of system operation from system testing under different given conditions. Conclusions will be drawn in the fifth section.

System Architecture

The Figure 1 presents the proposed scheme of hybrid transmission system, which aims to utilize the strengths of both panoramic video and regular video through layered transmission. The primary goal of panoramic system is to provide clear information on audio and video stream surveillance without blind spots, and enable people to check the scenes from the view of robot through convenient angle change on the client side. When real-time interaction is needed, video stream may be changed over to regular video stream of low latency to check any specific angle, so as to realize the real-time interaction and obtain remote videos with low bandwidth and low latency.

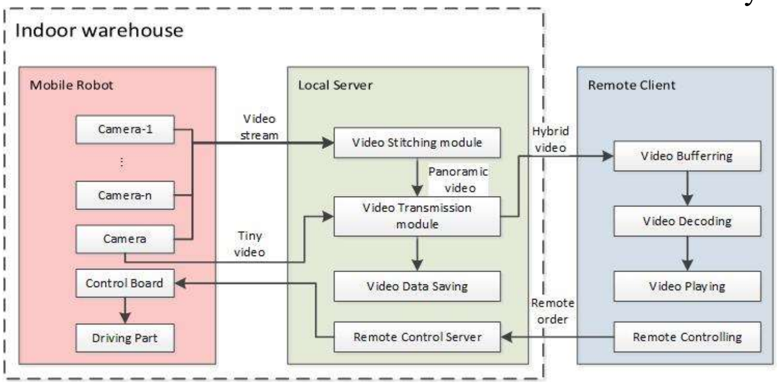

Fig. 1 Two-layer transmission scheme system architecture

This system is roughly divided into three parts, i.e. video capture control system of mobile robot, transmission and processing system of local server, and surveillance operation of remote client. Similar with regular video stream architecture, this system consists of video capture encoding, server transmission, and client decoding & playing. Mobile robot is armed with multiple cameras to capture the audio and video information of the environment in a timely manner, and encodes them in the standard compression format, e.g. H.264/AVC or HEVC/H.265. With wireless transmission system, streams are sent to local server, where they are further processed. The tiny video streams needed for interaction are distributed in the strategy of low latency, while the other video streams are stitched into 360-degree panoramic video for data storage and distributed transmission. The client side decodes and plays the streams, and supports the rendering of panoramic video and remote control logics, so the user may remotely watch the fetched clear panoramic video as needed, or switch to the tiny video for remote robot control. As it relies on a set of multi-camera video capture devices for high amount of data in streams and real-time interaction, this system can expand the visual system of existing robot system more easily and utilize streams and bandwidth more efficiently than the way of capturing through multiple sets of camera.

Detailed Module Design

Panoramic Video

Panoramic video is a new type of video format. As an image-based rendering (IBR) method, it can be represented in a plenoptic function model, which indicates the intensity of the light at the time t for wavelength from each angle of view (θ, φ) and the position of each camera.

$$p = P(\theta, \varphi, V_x, V_y, V_z, \lambda, t) \tag{1}$$

For simplification, the wavelength parameter (which may be parameterized by three major color values) and time parameter are removed to obtain the simplified function as follows:

$$p = P(\theta, \varphi, V_x, V_y, V_z) \tag{2}$$

If the position of camera is fixed, 360-degree panoramic video is therefore gained, which is the image of the surroundings around the camera from all angles.

$$p_{360} = P(\theta, \varphi) \tag{3}$$

The plenoptic function of panoramic video can be mapped onto a two-dimensional plane by corresponding the angle (θ, φ) to the coordinates x, y. The Figure2 presents a mapped 4K resolution panoramic image.

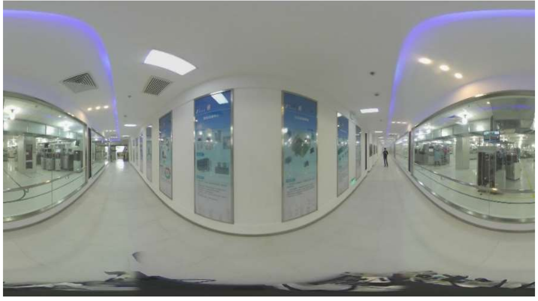

Fig. 2 Example Panoramic 360 degrees image

A camera may be moving, so plenoptic function is scarcely distributed in the dimensions Vx, Vy, Vz . The formula is as follows:

$$p_{360} = P(\theta, \varphi, V_x(t), V_y(t), V_z(t)) \tag{4}$$

Hence, continuous 2D images are mapped into the 3D space along the elapse of the time t, so as to form the shown panoramic video.

This 360-degree video without blind spots is very suitable for surveillance of mobile robot. With this technique, robot can capture the visual data via multiple cameras, which can be stitched at the server for storage and transmission of panoramic stream.

Video Frame Stitching.360-degree panoramic video relies on real-time stitching of multiple video streams. Stitching parameters are set based on the position of each camera to control its mapping position. The mapping relationship is presented in the following Figure3.

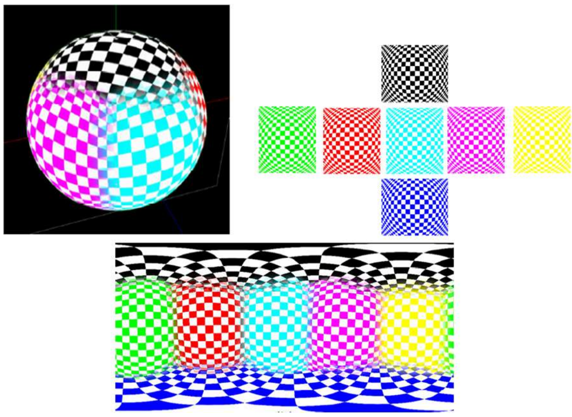

Fig. 3 Panoramic video stitching mapping

Nevertheless, massive hardware calculations must be conducted for stitching multiple video streams, which often leads to high latency of video streaming. This system can achieve better real-time function by transmitting a portion of video streams separately to avoid the latency caused by stitching.

The server can perform the real-time stitching of multiple video streams through video stitching software, encode and compress the generated 360-degree panoramic video for storage and transmission, so as to provide media data for remote client. The processing flow on the server side is presented in the Figure4.

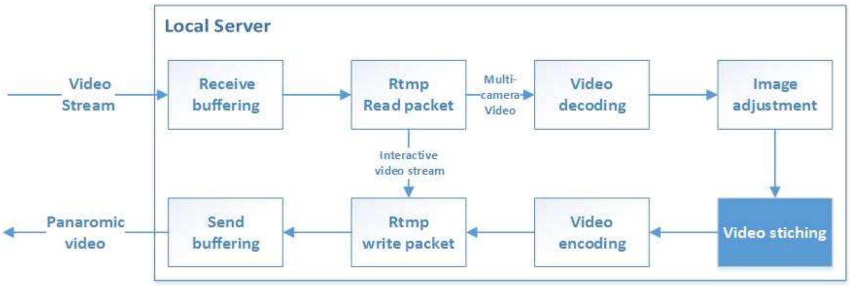

Fig.4 Server side video processing flow

Client Side Techniques. The client side allows to watch 360-degree panoramic video and remote control streams, and perform remote control and interaction. The common functions of client side include data buffering, data decoding and playing, but panoramic video requires mapping the images of video as a layer in a virtual sphere, which allows a user to be immersed in the space if his visual angle is at the center of the sphere.

The key point and coordinate mapping function of the sphere is as follows:

$$x = \cos(2\pi \times s \times S) \times \sin(\pi \times r \times R)$$

$$y = -\sin(-\pi \div 2 + \pi \times r \times R)$$

$$z = \sin(2 \times \pi \times s \times S) \times \sin(\pi \times r \times R) \tag{5}$$

In which, R and S stand for the reciprocals of image partitions, while r and s are the x,y coordinates in the image.

The client side is developed by smart device based on Android operating system, which can conveniently switch to the panoramic video through its built-in gyroscope for supervisor during surveillance. Meanwhile, the client side features remote control command and real-time interactive video stream change, so it can be changed over to real-time streams during remote control and interactive operation, and operate the robot on a dedicated interface.

Video Frame Compression

To transmit high-resolution video frame data through wireless connection, robot must compress the streams on its stream side before. Panoramic video requires very high resolution (normally 4K or above) to support the broad angle of view, so it needs much higher code ratio than regular videos, and highly efficient encoding algorithm for video frame compression. While capturing video data through multiple cameras, mobile robot encodes and compresses the video streams under the standards H.264/AVC and HEVC/H.265. H.264/AVC is an encoding technique based on interframe compression, which can achieve efficient prediction and encoding based on I-frame through intra-coded picture (I-frame), predicted picture (P-frame) and bi-predictive picture (B-frame) as well as high compression ratio. During transmission of compressed video streams in a noisy wireless channel, any loss of encoded frame will also have noticeable visual influence on the neighboring frames. Additionally, P-frame often causes noticeable latency at the time of decoding due to the interframe reliance of video, which can be changed by adjusting the interval in group of pictures (GOP). This encoding algorithm requires certain capacity of computation, so it normally encodes the video streams of high resolution in a real-time manner by means of hardware encoding.

In this system, stream collection side is equipped with the device supporting the outputs of hardware encoding, and each output is the 1080P flv stream encoded as h.264. As stitched video streams have higher resolution and lower requirements for latency, local server side employs the encoding frame based on FFMPEG, and encodes the video through CPU operation of the server to output the encoded 4K (4096x2048) data streams.

Video Frame Transmission

Data streams can be transmitted in two ways, i.e. sequential stream transmission and real-time stream transmission.

Sequential stream transmission downloads the files into the temporary folder of local client system via HTTP protocol, and plays the downloaded files in the player. It is featured by low pressure of transmission server and easy distribution and expansion, but causes higher latency of video streaming and inability of dynamic interaction. Real-time stream transmission relies on specific stream server to distribute and transmit the videos, so it offers lower latency and good interaction, but its data transmission is supported by real-time stream transmission protocol. Real-time stream transmission protocol belongs to the application layer protocol in the internet TCP/IP five-layer system architecture, including common protocols RTP, RTSP and RTMP.

The video data collected by robot should be transmitted to local server through transmission protocol, which relies on RTMP protocol for transmission and control of real-time video streams. RTMP is short for Real Time Messaging Protocol, a protocol proposed by Adobe to transmit audio & video and data. RTMP protocol effectively guarantees the quality of media transmission to present high-quality multimedia to users. Taking TCP protocol for its transmission layer, RTMP avoids the loss of quality due to packet loss during the transmission of multimedia in the wide area network (WAN). Additionally, the H.264 video encoding technique supported by FLV packaging format and based on RTMP protocol can keep good-quality images of transmitted video at very low code ratio, so as to guarantee the quality of transmitted video with low latency. Hence, it is very suitable for stream transmission with low network bandwidth.

The basic architecture of RTMP protocol data is described inFigure 5.

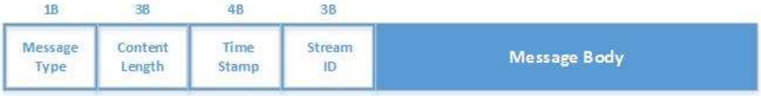

Fig. 5Rtmp message data structure

During transmission, message should be partitioned into the following chunks for transmission as Figure 6 presents.

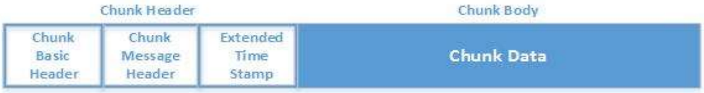

Fig. 6Rtmp chunk data structure

This system utilizes the librtmp library module, since library function can properly process the bottom logic of data processing. Several features of function in the process of transmission are as follows:

rtmp_create(const char* url) create rtmp dialogue via url address

rtmp_destroy(rtmp_trtmp) close and destroy rtmp dialogue

rtmp_handshake(rtmp_trtmp) rtmp handshake consists of three processes as follows:

rtmp_dns_resolve(rtmp_trtmp) resolve the dns of ip address

rtmp_connect_server(rtmp_trtmp) connect target server

rtmp_do_simple_handshake(rtmp_trtmp) do simple handshake

rtmp_play_stream(rtmp_trtmp) play stream

rtmp_publish_stream(rtmp_trtmp) publish stream

rtmp_read_packet(rtmp_trtmp, char* type, u_int32_t* timestamp, char** data, int* size) write packet into stream

rtmp_write_packet(rtmp_trtmp, char type, u_int32_t timestamp, char* data, int size) read packet from stream

Evaluation Experiment

The performance of video transmission system is evaluated by controlling the setting of parameters for different video streams. We deploy the indoor environment with each wireless node covering the radius of about 100m, and arrange a PC as the server. The indoor environment and mobile robot in this experiment are presented in the Figure7.

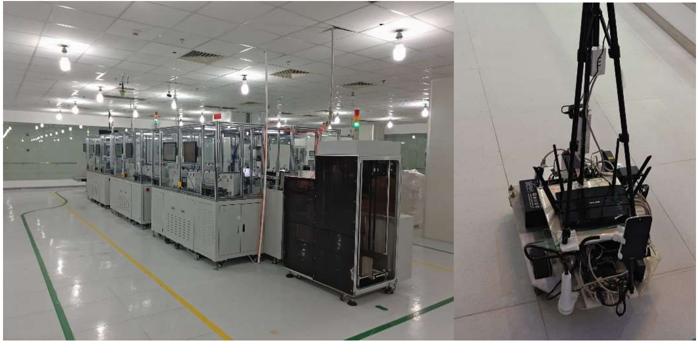

Fig.7.System testing environment and experiment mobile robot

This experiment intends to measure the performance of all elements in this system with different parameters. The controlled parameters of the system include video stream frame rate and the resolution of image, etc., while the tested performance indexes include video stream latency, frame loss rate, buffer size during video playing, etc. The control conditions include resolution and code ratio of streams and coverage of network. The panoramic video is compared with interactive video to evaluate the performance of the system as follows:

Video frame latency: the time interval from encoding of video data to their decoding and playing

Video frame loss rate: the percentage of video frame data successfully decoded

Player buffer size: the amount of valid data in the remote client buffer zone, which is used to analyze the stability of transmission.

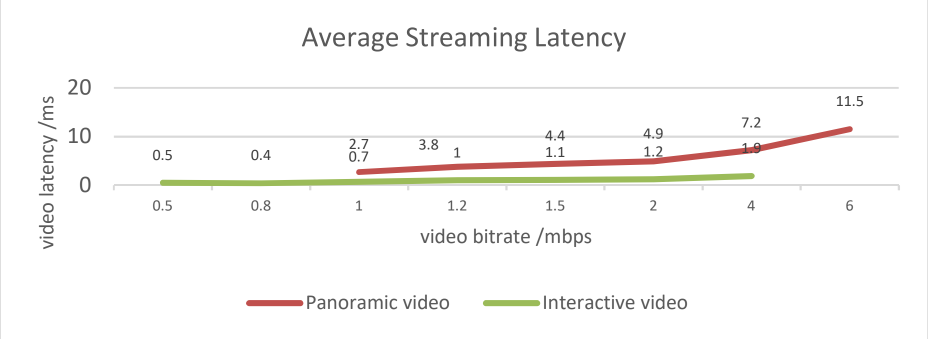

Fig.8 System steaming average streaming latency testing

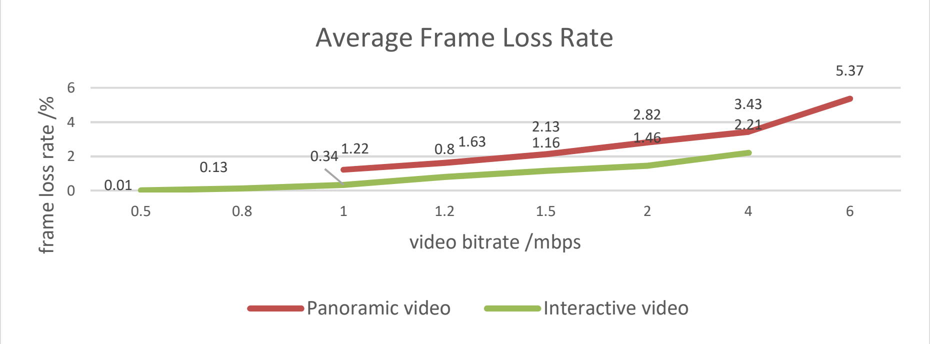

Fig.9 System steaming average frame loss rate testing

By configuring the streams at different code ratios, the operation results of the system are presented inFigure8 and Figure9. The stream transmission achieves good performance in the normal range of configuration. When the resolution is 720P(1280 x 960) and the data rate is 1mbps, the average streaming latency of interactive video is 0.7s, and its packet loss rate is about 0.34%. Hence, it achieves good performance in the remote control real-time interaction and other applications. For 4K resolution panoramic video, the average streaming latency is 7.2s and the packet loss rate is 3.43% when the code ratio of panoramic video is 4mbps. The clear panoramic video is mainly used for data storage and remote surveillance, so it is lowly sensitive to latency, which guarantee the stable remote transmission of data streams at an acceptable success rate of transmission. The results show that panoramic video offers good latency and frame loss rate when the code ratio is not changed, so it is more useful to switch channels for watching different video streams.

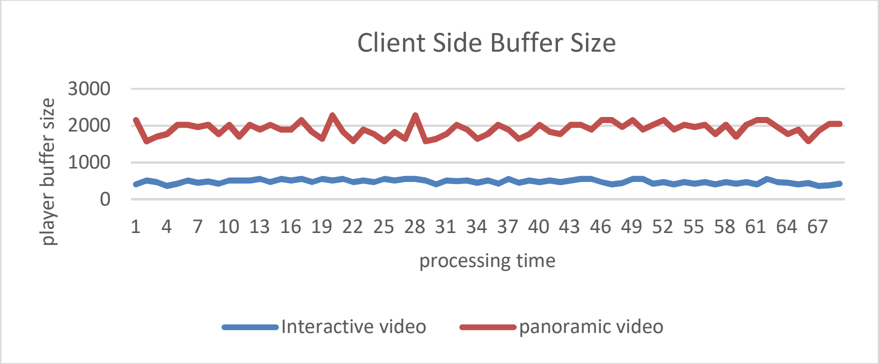

Fig.10 System steaming demo client-side buffer size

Client side buffer size reflects the quality of stream transmission and the experience of watching. The Figure10 presents the results of buffer size within a certain period when the system operates stably. The tested video streams are 4mbps panoramic video and 1mbps interactive video. As revealed in the figure, panoramic video needs larger buffer size, has a stable fluctuation of buffer size on the whole, receives the data stream packets stably, and can be encoded and played smoothly. The effect of playing is presented in the following Figure11.

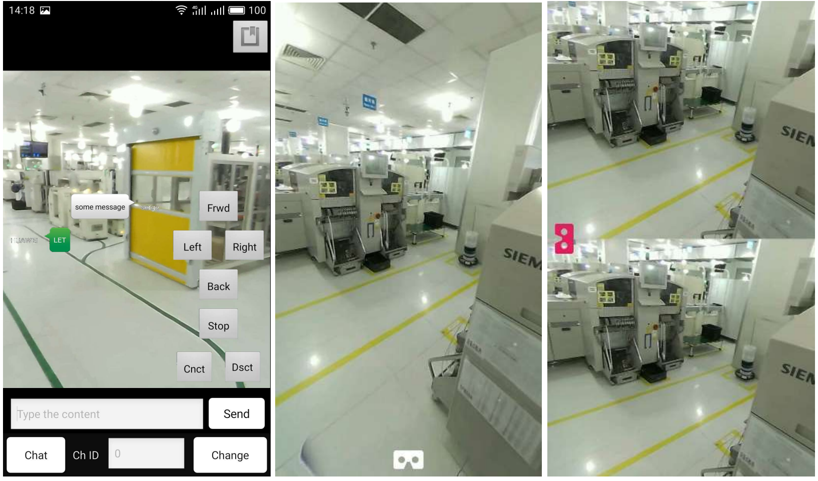

Fig. 11 Client control interface and panoramic video play

Conclusion

This paper introduces a real-time surveillance video stream system based on mobile robot. With the layered and separate transmission of videos, this system combines the high quality of 360-degree panoramic video with the timeliness of real-time interactive video, and better satisfies the robot surveillance system requirement for video data quality and interactive operation by separating the video streams on the collection side for transmission. We carry out an experiment of video system and obtain the results showing that the system has great stability and low latency for video streams at normal code ratio under the certain bandwidth. Therefore, the proposed system can be widely applied in the surveillance with indoor mobile robot.

References

[1] Shuai, Y., &Herfet, T. (2016, January). Improving user experience in low-latency adaptive streaming by stabilizing buffer dynamics. In 2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)(pp. 375-380). IEEE.

[2] Wu, J., Yang, J., Wu, X., & Chen, J. (2013, December). A low latency scheduling approach for high definition video streaming over heterogeneous wireless networks. In 2013 IEEE Global Communications Conference (GLOBECOM) (pp. 1723-1729). IEEE.

[3] Budagavi, M., Furton, J., Jin, G., Saxena, A., Wilkinson, J., & Dickerson, A. (2015, September). 360 degrees video coding using region adaptive smoothing. In Image Processing (ICIP), 2015 IEEE International Conference on (pp. 750-754). IEEE.

[4] Zhao, P., Li, J., Xi, J., & Gou, X. (2012, November). A mobile real-time video system using RTMP. In Computational Intelligence and Communication Networks (CICN), 2012 Fourth International Conference on (pp. 61-64). IEEE.

[5] Deng, R., & Liu, G. (2015, September). Modeling of H. 264/AVC based video transmission distortion over wireless network. In Image Processing (ICIP), 2015 IEEE International Conference on (pp. 2735-2740). IEEE.