Send Orders for Reprints to reprints@benthamscience.ae

The Open Electrical & Electronic Engineering Journal, **2014,** *8,* **397-403**

A Survey on Video Transmission Techniques

Chunmiao Yuan[1,*] and Qingyong Yang[2]

[1] *School of Computer Science & Software, Tianjin Polytechnic University, Tianjin, 300387, China*

[2] *Tianjin Tiandy Digital Technology Co., Ltd, Tianjin, 300384, China*

Abstract: Common video transmission system includes the following work like video collection and encoding of sending end, network transmission and reception, decoding and analysis of receiving end. This paper mainly focuses on elaborating and concluding principles and related techniques involved in video transmission system, such as rate control, transmission control mechanism, video retrieval, intelligent video analysis and error processing of video etc. In conclusion, it points out existing problems of video transmission techniques at present and future development direction of the researches.

Keywords: Video transmission, rate control, video analysis, video retrieval.

1. INTRODUCTION

As application of network multi-media increases gradually nowadays, network video transmission has become a quite important research task and been able to applied to various fields like security and protection monitoring, web TV and video conference. Usually, the process of video processing is shown as follows: firstly, the sending end carries out video collection and encoding for related videos; then, videos are transmitted by network; finally, decoding and related video analysis are implemented at the receiving end. Work of the sending end involves video collection, encoding and decoding. Transmission control refers to data flow control and congestion controls which are performed because there are network bandwidth fluctuations, when video data is transmitted over the network. After the receiving end has received video data, it will implement playback or do related work like video retrieval and video analysis further according to users' specific demands. Next, this paper will summarize relevant techniques involved in video transmission and processing.

2. RATE CONTROL OF VIDEOS

A. Overview

In video transmission system where channel rate is limited, it is necessary to put output code stream of a video encoder in a sending buffer first and then transmit it to the receiving end by channels [1]. If code rate is larger than channel rate when the video encoder outputs code stream, the code stream will be accumulated constantly in the sending buffer, which will result in increase of communication delay. On the contrary, if output rate of the encoder is smaller than channel rate, the sending buffer will keep an idle state basically; as a result, small communication delay can be realized but it does not make full use of channel resources. Therefore, rate control technique is a key technology in video communication, which enables output rate of a video encoder to be matched with channel rate. It not only makes full use of channel resources but also reduces communication delay from one end to another so that quality of video communication is improved.

Rate control of video encoding is a very important problem in video transmission [2]. Basic principle of rate control (RC) algorithm is that it tries to adjust variation of code rate, which is results in changes in images, in order to make them adapt to limited channel bandwidth under the condition that stable image quality is ensured.

B. Common RC Algorithms

Application of rate-distortion (R-D) technique to rate control can prevent the buffer from overflowing or avoid its underflow under limited bandwidth as well as obtain optimum video quality. Thus, estimation or modeling of R-D is a key problem that the video encoder must solve. R-D can be described by using distortion-quantization (D-Q) function or rate-quantization (R-Q) function of the video encoder [3]. The two functions are called by a joint name, i.e., R-D function. Currently, models that are commonly used in the fields of video RC and RDO include classical RC model, quadratic model, logarithmic model, ρ domain model and mixed model etc [4].

Classical RC Model

In the classical RC model, it is assumed that image distortion probability accords with even distribution and then variance is used to evaluate image distortion. The image dis-

1874-1290/14

tortion in this model is only related to quantization parameter, as shown in (1).

$$D = \frac{Q^2}{12} \tag{1}$$

ρ Model

Literature [5] utilizes quantization step X to quantize the coefficient after a macro block has accepted DCT conversion. Proportion of bit '0' in the obtained result is defined as ρ. *Via* a number of experiments, the author finds that the code rate after encoding and ρ follows a linear relation, approximately. Therefore, the RC problem can be solved by establishing the relationship between code rate and ρ and the relation between ρ and QP. ρ can be solved by the following formula.

$$\rho(X) = \sum_{|a|<X} \rho(a) \tag{2}$$

Quadratic R-D Model

By virtue of researches, Chiang et al. [6] put forward a quadratic video R-D model in 1997. Firstly, it is assumed that information source accords with Laplace distribution, code rate is R and distortion is D. The logarithmic form in (3) can be used to express the R-D model:

$$R(D) = \begin{cases} \ln[1/(\alpha D)], & 0 < D < 1/\alpha \\ 0, & D \geq 1/\alpha \end{cases} \tag{3}$$

Where α is Laplace distribution parameter and α>0. If we use Taylor to develop the foregoing formula, adopt quantization step to measure distortion and assume that R and Q are encoding bit number and quantization step, respectively, the R-D function about RC in the logarithmic form in the foregoing formula will be converted into an expression form of quadratic polynomial, as shown in (4). The model is applied to MPEG-4 standard and obtains favorable RC effect.

$$R = a_1 \times Q^{-1} + a_2 \times Q^{-2} \tag{4}$$

In the foregoing formula, a_1 and a_2 are model parameters.

However, Lee et al. [7] deem that texture information and non-texture information should be distinguished when image coding is carried out. However, the quadratic model does not consider texture information. If mean absolute difference (MAD) of signals is defined as frame difference between predicted image and the original one, content complexity of images can be measured by MAD. In doing so, the quadratic R-D model that is related to content complexity of images and describes texture information about encoding, can be obtained, as shown in the following:

$$\frac{R_T}{MAD} = a_1 \times Q^{-1} + a_2 \times Q^{-2} \tag{5}$$

The model parameters a_1 and a_2 can be obtained by linear regression technique. When the parameters a_1 and a_2 are updated, ways to select sample data can be improved by prediction mechanism of encoding.

Linear Model

$$R = a\frac{MAD}{Q} + b \tag{6}$$

The model parameters a and b can be obtained by linear regression technique. It can be found that the linear model is simpler compared with the quadratic model. Besides, the quadratic model has extremal problems, so it cannot reflect monotonicity of a rate model completely.

Heuristic Distortion Model

Because changes in video images have spatiotemporal continuity, image distortion also has some spatiotemporal continuity in distribution between adjacent image frames. In view of this, Jiang *et al.* [8] propose a heuristic distortion model which not only considers spatiotemporal continuity of images but also takes the relationship between image distortion and quantization step into consideration.

$$D^k = x^k Q^k, \quad \text{‘’†} \quad x^k = \frac{D^{k-1}}{Q^{k-1}} \tag{7}$$

3. VIDEO TRANSMISSION CONTROL

A. Introduction

As we all know, the current Internet only provides 'best effort' service at network layer to simplify network implementation, i.e., it is only responsible for sending in network rather than paying attention to whether packets can arrive accurately according to related order within specified time or packets are lost or not, that is, it cannot provide quality of service (QoS) for packet transmission [9]. This transmission mechanism is appropriate for abrupt data transmission in network [10].

Aiming at the foregoing problems, researchers propose different video transmission plans, such as classification method [11, 12]: it classifies code stream and transmits different types of video data dynamically according to current status of network. Supplementary information method [13, 14]: when data are sent, extraneous information needed by encoding and decoding is also sent to solve the QoS problem of video transmission. Environmental adaptation method [15, 16]: according to different network environment, different transmission modes are utilized to carry out video transmission; and data partitioning method [17, 18]: data are partitioned according to their features and then transmitted.

B. Common Video Transmission Control Schemes

When video transmission is performed, transmission control schemes of video transmission can be divided into the following four types according to different control schemes about implementation of video transmission.

RC Based on the Sending End

The method implements RC at the sending end [7] which will monitor current status of the network and adjust sending rate according to network status in time.

RC Based on the Receiving End

In this method, the receiver adjusts rate [19], while the sender does not participate in RC.

Traffic Shaping

Traffic shaping [19] is also called packet shaping, which is a management technology about computer network flow and delay packet transmission to adapt to network status.

Mixed-type RC

Mixed-type RC is formed by combining RC based on the sending end and RC based on the receiving end.

4. INTELLIGENT ANALYSIS OF VIDEOS

A. Introduction

When traditional video analysis system does work like analysis and recognition of video content, such work needs to be realized by people. Since its workload and labor intensity are large and faults may be caused easily when analysis are carried out artificially, intelligentization is becoming a development direction of video analysis. One of the most important problems of the field of video intelligent analysis is how to find and analyze useful information automatically from a large number of video images [20]. It is essential for the video intelligent analysis to have the following functions, such as recognizing objects automatically, perceiving changes in target objects and having the ability to realize timely and alarm automatically when abnormal conditions appear.

B. Common Video Intelligent Analysis Techniques

Target Detection

Target detection extracts regions of variation from background images in continuous pictures. Partitioning moving regions effectively is the most important step in the process of target detection, which will affect post-processing of target objects, such as classification and behavior identification. Target detection involves [21] that moving targets are detected in different images that have accepted binarization. Several common dynamic video target detection methods contain background subtraction, time difference method and optical flow method [22].

Target Tracking

Target tracking technique [23] is a process where one or several moving targets are located in videos within a period of time. Target tracking need support of target recognition program, and is a time-consuming process. Target tracking is an important direction in the field of computer technology and has been widely applied to many fields like science and technology, national defense construction, aviation and space navigation, medicine and health and national economy.

Selecting the right features plays a critical role in tracking [24]. Many tracking algorithms use a combination of these features. The details of common visual features are as follows.

Color: The apparent color of an object is influenced primarily by two physical factors of the spectral power distribution of the illuminant and the surface reflectance properties of the object.

Edge:. Object boundaries usually generate strong changes in image intensities. Edge detection is used to identify these changes.

Optical Flow: Optical flow is a dense field of displacement vectors which defines the translation of each pixel in a region.

Texture: Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity.

By partition targets reasonably, extract features accurately and perform matching of feature information, we can recognize targets correctly. At the same time, run time of algorithm must be considered in order to ensure real-time performance. When moving speed of a target exceeds frame rate or the target constantly changes its direction, target tracking will have some difficulty. The main challenge that target tracking faces with is that uncertainty of environment results in the situation that it is difficult to estimate status of targets. Additionally, factors like changes in illumination and shapes of targets, complicated background and covering affects robustness of tracking. In order to overcome such difficulty, researchers put forward many target tracking methods. Common target tracking algorithms [25] include mean shift, Kalman filtering and particle filter etc.

Mean shift is a nonparametric clustering technique, which neither needs priori knowledge about quantity of clusters nor restricts shapes of clusters and it acts as an iterative algorithm, as well. Firstly, it computes mean shifting value of the current point and moves the point to shifting mean vector. Then, it uses the points as a new starting point, continues the foregoing process and ends the iterative process until some conditions are satisfied. It was in 1975 that the method was proposed for the first time. Since then, it has been widely studied in the field of video analysis and huge progress has been achieved.

Mean shift algorithm is a nonparametric fast mode matching algorithm based on kernel density estimation. With respect of variable a the finite data set in Euclidean n-spaces, sample average near the point x (x X) is defined as:

$$sm(x) = \frac{\sum_{a} K(a-x)\_(a)a}{\sum_{a} K(a-x)\_(a)} \qquad a \in A \tag{8}$$

In the foregoing formula, K is a kernel function, ω stands for weight function and the difference value $sm(x)-x$ is called mean shift vector. Mean shift algorithm moves data points towards direction of the mean shift vector repeatedly and constantly until convergence is eventually obtained. Movement step is a fixed value. At this time, position of kernel center is corresponding to extreme value of a certain probability density, i.e., position of schema.

Basic thought of Kalman filter is that it does not consider impacts of input signal and observation noise first but obtains estimated values of state variable and output signal and corrects the estimated value of the state variable after the evaluated error of the output signal has been used for weighting to make mean square error of evaluated error of the state variable be minimum.

Particle filter is a filtering algorithm based on Bayesian reasoning. It is realized by non-parametric sequential Monte Carlo simulation method. Besides, particle filter is not restricted by the situation that system should satisfy linearity and Gaussian distribution. Thus, it has been widely applied to many aspects like navigation, machine vision and target tracking. In the particle filter, many features can be used. Color and edge contour [26] are two common features that express targets.

The main difference between particle filter algorithm and the mean shift algorithm is that the way in which they look for extreme value of similarity & probability density distribution is different. In detail, the mean shift look for position of extreme value by iteration, while the particle filter does not adopt the iteration process but spreads a number of particles by which it obtains position of extreme value. In addition, since particle sampling is spread to the whole similarity & probability density space, it has global optimum nature. The mean shift that is used for tracking belongs to deterministic algorithm, and the particle filter belongs to a statistical method. Compared with the particle filter, the mean shift tracking algorithm has better real-time performance. However, the particle filter holds better tracing accuracy theoretically.

Video Retrieval

The content-based video retrieval carries out automatic analysis of video content according to its features and makes index in order that users can find images or video clips with specific features or specific content in database. The difference between it and traditional retrieval is that the latter is based on key words of text and the content-based video retrieval is on the basis of visual features. In practical application, video content cannot be described clearly by using several key words or a paragraph of simple characters. Usually, users hope system can find broadcast video clips automatically as long as retrieval examples or feature description is given, for example, the content-based video retrieval can be realized [27].

The content-based video retrieval can find corresponding matched images or video clips from the video database according to information like color, texture and shape of each frame of images in order to satisfy demands of retrieval. It carries out retrieval according to semantics and visual features of media targets. Besides, the process of retrieval is not keyword matching but matching among visual features so that it has many advantages like objectiveness, manpower saving, ability to establish complicated description and good universality.

There are four main processes [28] involved in content-based video indexing and retrieval: video content analysis, video structure parsing, summarization or abstraction, and indexing. Video content analysis is to map extractable visual features (such as color, texture, shape, structure, layout, and motion) into semantic concepts (such as indoor and outdoor, people, or car-racing scenes).Video structure parsing segments the video into individual scenes. Video summarization is the process of creating a presentation of visual information about the structure of video, which should be much shorter than the original video. Video indexing is retrieving or searching video databases by keywords or phrases.

Intelligent Behavioral Analysis

Human activity analysis is receiving increasing attention from computer vision researchers. This interest is motivated by a wide spectrum of applications, such as surveillance, man-machine interfaces, video coding, and so on. It has been a hot research topic in image analysis, psychology, and neurophysiology.

Compared with object identification and the tracking technique, action recognition technology is at a higher research level. At present, although some action details in video intelligent analysis system cannot be recognized accurately, analysis of objects' overall behaviors have been achieved, such as track, shape, color, running speed, velocity change and volume of objects. By analyzing these data, some demands for features in the field of video application, such as alarm about restricted zone, quantitative statistics, medical monitoring and intelligent conference, can be realized. Intelligent behavioral analysis involves extraction and expression of motion features at the bottom layer, representing methods of behaviors and inference patterns of behavioral semantics on the higher layer.

Image representations can be divided into two categories [29]: global representations and local representations. Local representations describe the observation as a collection of independent patches. The calculation of local representations proceeds in a bottom-up fashion: first, we will detect spatio-temporal interest points and calculated local patches around these points, then we combine the patches into a final representation. Global representations encode the Region Of Interest (ROI) of a person as a whole. The ROI is usually obtained through background subtraction or tracking. Common global representations are derived from silhouettes, edges or optical flow.

Extraction and expression of vision or other motion features at the bottom layer are basic work needed by implementation of inference about behavioral comprehension on the higher layer. Common ways to express moving targets

include static characteristic based on edge, dynamic characteristic based on light stream and spatiotemporal feature obtained from some parts. The method based on spatiotemporal interest point can provide abundant description and expression. In this method [30], spatiotemporal words are composed of extracted spatiotemporal interest points for each segment of video sequence. In complicated videos, they can be expressed by response function of linear filter.

(9)

Where I stands for image; $g(x,y,\sigma)$ is two-dimensional Gaussian smoothing kernel, which is appropriate for spatial dimension (x, y); *hev* and *hod* are one-dimensional orthometric filter applied to time dimension.

When action recognition is carried out, we usually match recognized image with behavioral reference sequence set is selected in advanced in order to judge specific behaviors. It is a typical classification problem. Action recognition can be realized by two methods, i.e., state-space method and template matching [31]. The state-space model defines each static posture as a stage, i.e., these states are connected by a certain probability. When the state-space model carries out action recognition, it implements traversal for all of these different states and then chooses maximum joint probability as the result of recognition. The most representative state-space model is Hidden Markov Model (HMM) [32]. HMM can process time-varying data effectively and has learning and recognition capacity. Template matching coverts representative images into a group of static templates in advance and then compare images with behavioral templates in the process of recognition. In addition, dynamic bayesian network (DBN), conditional random field (CRF) and decision-making tree can be applied to recognition.

Currently, researches on behavioral analysis usually only involve action recognition. Processed behaviors can be divided into two types. One type contains finite rule behaviors which are quite simple, such as sitting, walking, running, standing and crouching. The other type refers to finite behaviors processed in a specific scene, such as driving fatigue recognition and fighting. In this scene, behaviors have strict limitations and behavior description usually adopts speed or trail of movement. However, there is a huge gap between comprehension of behavioral analysis in a random scene and practical application.

5. ERROR PROCESSING OF VIDEOS

A. Introduction

When real-time video data are transmitted on network, the frame rate of correct decoding of the receiving end will be much lower than the sending frame rate of the sending end if video data frames are sent under the situation that bandwidth is not enough or packet loss probability is high for impacts of factors like bandwidth and packet loss probability. Two types of errors [33] may be caused when videos are transmitted: erasable errors and random order ones. Erasable errors are caused by burst error of packets in network, which are short-term system errors. Random order errors results from defects existing in physical channels (such as order insertion, deletion or reversal). Since most of video coding techniques adopt variable-length encoding (VLC) technique, the two types of errors do not have much difference f or video transmission. To realize detection of errors and error concealment, redundancy processing is often needed. Generally speaking, redundancy may exist in source encoder, entropy encoder and channel encoder.

B. Common Error Processing of Videos

In traditional video communication system, error correction technique, Automatic Repeat-reQuest (ARQ) and forward error correction (FEC) are usually used to ensure correctness of transferred information.

ARQ

ARQ is one of the error correction protocols of data link layer (DLL) in OSI model. By using the confirmation and overtime mechanisms, it realizes reliable information transmission based on unreliable service. Because of design mechanism of ARQ, it is not appropriate for video transmission. The reason for this is that data retransmission will affect real-time performance of video transmission. When network performance is not good, packets need be retransmitted frequently and excessive retransmission will worsen network performance and reduce handling capacity of channels. Besides, feedback mechanism of ARQ cannot be realized in application of broadcast and multicast multicasts.

FEC Mechanism

FEC is a widely applied error control technique when information is transmitted in unreliable or noisy channels in communication system, information field and encoding field [34]. Specifically, FEC is a data coding technique. After the sender has done data preprocessing and data has been transmitted to destination, the receiver will detect errors.

With respect to realization of FEC [35], it increases error correction information by FEC algorithm based on k original data packets which decodes data into n packets ($n>k$) in order that original k packets can be restored according to error correction information. As long as the receiving end receives data packets whose quantity is larger or equal to k. k/n is called code rate, which indicates the proportion of effective data to total data after encoding. Obviously, the smaller the value of k/n is, the stronger the error correcting capability is but the higher the rate of redundant data is. When n and k are determined, the probability at which all data can be received correctly can be calculated:

$$P(n,k) \approx \sum_{l=k}^{n} \binom{n}{l} (1-p)^{l} p^{n-l} \tag{10}$$

In FEC method, the receiving end can not only find errors but also determine the position where errors happen and correct them. The FEC method uses error correcting code and there is no need to inform the sending end to retransmit

when it finds errors. This is the difference between FEC and ARQ method. While FEC receives packets at the receiving end, it uses FEC decoding and opens such packets in order to obtain order stream of output and construct output video. At this time, it is no need to carry out data retransmission. Thus, FEC technique is appropriate for real-time video transmission and can be applied to video monitoring system. Advantage of FEC is that it has low processing cost and accurate error recovery, while its disadvantage is that transmission of redundant data used for error correction will cause waste for bandwidth.

Video transmission can perform different FEC error correction according to different strategies, such as FEC layering QoS strategy [36], self-adapting FEC algorithm based on video layering and tree-based self-adapting FEC algorithm.

Error Resilience

Generally speaking, transmission error will not result in unacceptable distortion when videos are reconstructed. Since error resilience mechanism needs more redundant data to recover data, efficiency is low. The reason for this is that it needs to transmit more bits to obtain the video effect with the same quality. Actually, this is data redundancy. The error resilience mechanism of videos directs at obtaining maximum benefits by adding minimum redundancy.

Error Concealment

Error concealment method means that the decoder can estimate lost samples according to surrounding sample data of the received sample after a certain sample of video images has been lost since video sample space has some spatiotemporal continuity [37]. Being different the error resilience mechanism, the error concealment method does not introduce additional data to carry out error resilience but uses surrounding data of false images to recover data of such false images according to the smoothness of video images. However, it increases computation complexity of system.

Error concealment utilizes relevancy in time domain and space domain. In another word, error concealment itself will not change error codes but takes some post-processing techniques when the receiving end carries out decoding and makes every effort to recover damaged images in order so these error codes cannot be perceived by eyes visually. Considering the problem, i.e., error codes of video transmission, it can be solved by using error concealment technique. The error concealment technique can be implemented in the time domain, i.e., time concealment method, or space domain, i.e., space concealment method. The error concealment technique of time domain usually adopts characteristics of predictive coding according to video compression. In this technique, images that are adjacent in time can be used to recover false images. The space-domain concealment utilizes neighboring correct pieces to carry out pixel linear interpolation to recover lost pieces, which is a post-processing error concealment technique. The space-domain error concealment can be realized by linear weighting interpolation or region segmentation.

The challenges in error concealment are as following.

Predictive Coding: Due to spatial and temporal prediction, a single wrong recovered sample can lead to more errors in the same and following frames.

Variable Length Coding: Due to the nature of VLC, a single bit error can cause the decoder to lose synchronization.

Time Varying: Video source and network conditions are typically time-varying, so it is hard (almost impossible) to derive optimal solution based on statistical models (of source and/or channel).

Complexity: Video sources have very high bit rate, high complexity operations are not acceptable for real-time applications.

CONCLUSION

In recent years, video transmission has become hot application in Internet. This paper not only elaborates RC problem appearing in video transmission, transmission control scheme, intelligent analysis technique and video error processing technique but also carries out analytical comparison. Basic thought of RC algorithm is that it tries to adjust variation of code rate, which is caused by changes in images, in order to make them adapt to limited channel bandwidth under the condition that stable image quality is ensured. Transmission control technique focuses on how to ensure transmission quality of video in network. Intelligent analysis of videos, which is a research interest at present, uses algorithm to realize target location of video, tracking, behavioral analysis and related intelligent video monitoring. Error processing technique of videos aims at how to find and analyze errors generated when videos are transmitted in order to provide better visual effect. At present, as researches that have appeared in recent years, problems about behavioral comprehension of intelligent analysis field in video transmission processing system are drawing more and more attention.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is funded by the project of National Development and Reform Commission: 'A Project on Digital Security and Protection Technique and Core Products' Industrialization Base Construction' and the special project of Internet of Things of Ministry of Industry and Information Technology of the People's Republic of China" 'Research and Development of Several Key Sensors and Network Techniques in Internet of Things for Disaster Reduction and Prevention & Emergency'.

REFERENCES

[1] M-li Chen, Peng Liu, Q.-D Yao, "Macroblock Layer Rate Control Algorithm of H.264/AVC", *Journal of Zhejiang University (Engineering Science)*, vol. 41, no. 6, pp. 925-929, 2007.

[2] D. Wu, Y. T. Hou and Y. Q. Zhang, "Transporting Real-timing Video over the Internet: Challenges and Approaches", In: *Proceeding of the IEEE*, vol. 88, no. 12, pp. 1855-1875 , 2000.

[3] T. Chiang and Y. Zhang, "A New rate control scheme using quadratic rate distortion model", *IEEE Transactions on Circuits System on Video Technology*., vol. 7, no. 1, pp. 246-250, 1997.

[4] M. Jiang, X. Yi, and N. Ling, "Improved Frame Layer Rate Control for H. 264 using MAD ratio", In: *Proceedings of the 2004 international symposium on circuits and systems*, pp. 813-816, 2004.

[5] Y. Wu, S. Lin, and Z. Niu, "Efficient rate control schemes for H.264/AVC", *Chinese Journal of Computers*., vol. 31, no. 2, pp. 329-339, 2008.

[6] L. Xu, W. Gao, and X. Ji, "A rate control algorithm for scalable video coding", *Chinese Journal of Computers*., vol. 31, no. 7, pp. 1-10, 2008.

[7] H. J. Lee, L. Chiang, Y. Q. Zhang, "Scalable rate control for MPEG-4 video," *IEEE Transactions on Circuits and Systems for Video Technology*., vol. 10, no. 6, pp. 878-894, 2000.

[8] Z. Z. Li, and R. X. Kai, "Research on the Improvement Method of Rate Control Based on H. 264", *Journal of Zhejiang University. Technology*., vol. 36, no. 5, pp. 519-522, 2008.

[9] Q. Zhang, W. Zhu and Y. Q. hang, "End-to-end QoS for video delivery over wireless internet", *In: Proceedings of the IEEE*, vol. 93, no. 1, pp. 123-134, 2004.

[10] C.-K. Jin, W. Jia, and S. L. An, "End-to-end congestion control strategy for streaming media based on TCPW", *Journal of Image and Graphic*, vo1. 13, no. 10, pp. 1959-1963, Oct. 2008.

[11] S. Xiao, C. Wu, and Y. Zhou, "Reliable visual transmission method combining source statistics and network congestion control", *Journal of Software*. vo1. 18, no. 11, pp. 2882-2892 2007.

[12] M. K. Jubran, M. Bansal, and L. P. Kondi, "Accurate distortion estimation and optimal bandwidth allocation for scalable H.264 video transmission over MIMO systems" *IEEE Transactions on Image Processing*, vol. 18, no. 1, pp.106-116, 2009.

[13] J. Wang, A. Majumdar, and K. Ramchandran, "Robust video transmission with distributed source coded auxiliary channel", *IEEE Transactions on Image Processing*., vol. 18, no. 12, pp. 2695-2705, 2009.

[14] V. K. Goyal, "Multiple description coding: compression meets the network", *IEEE Signal Processing Magazine*., vol. 18, no. 5, pp. 74-93, 2001.

[15] D. Hu, F. Liu, and X. Zhu, "Implementation of a prototype real-time multi-mode wireless video transmission system", *Journal on Communications*, vol. 27, no. l0, pp. 107-113, 2006.

[16] Y. Ma, Y. Chang, and J. Yao, "Novel transmission method for H.264/AVC video data over wireless OFDM channels", *Journal of Xidian Universityersity*, vol. 35, no. 3, pp. 396-405, 2008.

[17] C. Hou, C. Ni, and Y. Hou, "Unequal error protection transmission scheme for H.264 video streams based on EWF", *Journal of Tianjin Universityersity*., vol. 25, no. 4, pp. 309-314, 2012.

[18] Y. Zhang, and K.-K. Ma, "Error concealment for video transmission with dual multiscale markov random field modeling", *IEEE Transactions on Image Processing*., vol. 12, no. 2, pp. 236-242, 2003.

[19] Z. G. Li, F. Pan, and K. P. Lim, "*Adaptive basic unit layer rate control for JVT*", JVT of ISO/IEC MPEG&ITU-T VCEG, JVT-G012-rl.doc, 2003.

[20] S. H. Ren, S. H. Wang, and S. L. Liu, "Wireless Videos Surveillance System Based on H.264 and 3G technologyogy and optimized rate control algorithm", *Application Research of Computers*., vol. 27, no. 4, pp. 1554-1556, 2010.

[21] M. Valera, and S. A. Velastin, "Intelligent distributed surveillance systems: a review", *In IEEE Proc.-Visual Image Signal Process.*, vol. 152, no. 2, pp. 192-204, 2005.

[22] R. Brunelli, and T. Poggio, "Face recognition: features *vs.* template", *IEEE Transaction on Pattern Analysis and Machine Intelligence*., vol. 15, no. 10, pp. 1042-1052, 1993.

[23] S. Tekinalp, and A. A. Alatan, "Utilization of texture, contrast and color homogeneity for detecting and recognizing text from video frames", *In: Proceedings of International Conference on Image Processing.*, 2003, 2: III_505-508.

[24] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey". *ACM Computational Survey*., vol. 38, no. 4, Article 13 December 2006. 45 pages. [Online] Available: http://doi.acm.org/10.1145/117-7352.1177355

[25] Yilmaz, "Object Tracking: a Survey", *ACM Computational Survey*, vol. 38, no. 4, pp. l-45, 2006.

[26] K. A. Bashirfi, and D. Schonfeld, "Object trajectory based activity classification and recognition using hidden markov models", *IEEE Transactions on Image Processing*., vol. 16, no. 7, pp. 1912-1919, 2007.

[27] JOURNAL Zhang, "Robust shot boundary detection and video summarization based on motion information", *Journal of Computer-Aided Design Computer Graphics.* vol.22, no.6, pp. 1023-1032, 2010.

[28] N. Sebe, M. S. Lew, and A. W. M. Smeulders, *Video Retrieval and Summarization*, Computer Vision and Image Understanding, 2003, (in press).

[29] R. Poppe, "A survey on vision-based human action recognition", *Image and Vision Computing*, vol. 28, pp. 976-990, 2010.

[30] Z. Zhang, Y. Hu, and S. Chan, " Motion context: a new representation for human action recognition", In: *Proceedings of the European conference on computer vision (ECCV'08) – part 4, lecture notes in computer science*, Marseille, France, October 2008, pp. 817-829.

[31] N. Habili, C. L. Cheng, and A. Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 8, pp. 1086-1097, 2004.

[32] G.-Y. Zhu, and C.-S. Xu. "Action recognition in broadcast tennis video", In: *IEEE Proceeding of the 18th International Conference on Pattern Recognition*. New York: 2006, pp. 251-254.

[33] Y.-L. Gan and W.-L. Zhong, "Error concealment techniques in video communication", *Application Research of Computers* vol. 20, no. 2, pp. 8-9, 2003.

[34] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links", *IEEE Transactions on Wireless Communications*, vol. 3, no. 5, pp. 1746-1755, 2004.

[35] Z. Mei, and J-T. Li, "An adaptive forward error correction algorithm for streaming video", *Journal of Software*., vol. 15, no. 9, pp. 1405-1412, 2004.

[36] R. Wang, JL-M. Hu, and M.-Q. Jiang, "Design of a video surveillance system based on QoS strategy", *Microelectronics & Computer*., vol. 23, no. 6, pp. 68-71, 2006.

[37] Y. Wang, and Q. Zhu, "Error control and concealment for video communication: A review", In: *Proceedings of the IEEE*, vol. 86, no. 51, pp. 974-997, 1998.

Received: June 09, 2014 Revised: June 22, 2014 Accepted: July 24, 2014