Multi-regions in the Scalable Video Coding Method

Ce Li, Jianru Xue, Xuguang Lan, Miao Hui, Le Wang, Nanning Zheng
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
Xi'an, 710049, China
celi@aiar.xjtu.edu.cn

***Abstract*—Human visual system function is non-uniform to visual perception. Normally, people tend to focus on one or more specific moving objects while watching video. In addition, video transmission is flawed due to heterogeneous networks and limited bandwidth. Therefore, how to transmit the content of interests more efficiently becomes a growing concerning topic. This paper proposes a method of multi-region in the scalable video coding. In the proposed method, the multi-region firstly get through the particle tracking filters, and then it is implemented motion estimation to each region and finally allocated bit-rate under the 3D-wavelet scalable video coding model. Experimental results show that the performance of our approach is effective, in particularly the situations of lower bit rate. The approach is particularly suitable for mobile video communication, and video surveillance on intelligent vehicle.**

I. Introduction

The rapid development of internet video streaming technology will inevitably encounter the bottleneck of limited bandwidth. Low bit-rate video compression will cause loss of some detailed information, even the region which people are particularly interested in. Traditional video coding standard is facing a lot of new challenges, as it does not provide the flexibility to adapt various transmission conditions and the diversity of customer needs. In this paper, based on a solid foundation of motion compensation temporal filtering (MCTF), how to effectively remove the temporal redundancy in multi-region coding is explored, so as to enhance the reconstructed image quality at low bit-rate.

Currently scalable video coding technology has two core transforms: one is based on discrete cosine transform (DCT); the other is based on discrete wavelet transform (DWT). The extension of the H.264/AVC standard is based on DCT. It implements temporal, spacial and quality scalable joints by adding MCTF. Although the coding method complies with traditional standard video coding system, it still requires a higher computational complexity. The other is based on the DWT of the three-dimensional wavelet coding. It is flexible in representing non-stationary and has the ability to adapt to human visual characteristics. Through the time-domain high and low frequency frames embedded coding, it will be real-time, joint scalable in space and quality. In addition, since the JPEG2000 and MPEG-4 put forward the concepts of region of interest (ROI) and related algorithms, many researchers focus on what is ROI, and the content of scalable video coding in the core platforms, which are give their respective solutions.

According to human visual perception of non-uniformity, it becomes a concerning topic of current research that realizes multi-resolution for video content coding and transmits for the user concerned about the regional distribution of higher bit-rate technology. Tae et al. [1]proposed a method of description multi-region by using H.264/AVC flexible macro-block order (Flexible Macro-block Ordering, FMO) technique. Though it organizes a number of video macro-block layered packing regional transmission, the method can be scalable transmission method and adequate variation in the heterogeneous network. But it only applies to the core for the DCT transform in the framework of the H.264/AVC scalable coding technology, which could not be directly in support of DWT as the core transform realizes SVC system.

Similarly, Kim et al. [2] proposed another method of using the FMO of the H.264/AVC, which divided a video frame into different pieces of mobile phone screen rather than the size of ROI coding independent program. The solution of the proposal can solve a number of summary ROI overlap coding regions sharing problem. But, it is not given effective solutions that how to effectively transfer the video content of high concern for the network transmission bandwidth reduced. In particular, Chen et al. [3] gave a map to use significant ROI coefficient control region of the video bit-rate coding schemes by JSVM platform. This method gave video human face saliency map by a simulation goal-driven human visual system (Top-Down) and data-driven (Down-Top) model, in accordance with the region where significant frame rate coefficient ratio of the transmission functioned. Whereas, in this method, the saliency region is only the face region, and not fully realized by using regional motion estimation of ROI frame redundant in this method.

Generally, motion objects of video sequences are often the most concerned objects. Based on this hypothesis, the contribution of this paper is twofold. Firstly, this paper uses a new method to combine multi-target object tracking with ROI scalable video coding to produce an integrated system. Secondly, on the" T+2D " three-dimensional wavelet scalable video coding system platform, the paper presents a scalable video coding model to local multi-region motion estimation respectively, as well as lifting the bit-plane can be adapted to a variety of network transmission conditions. In addition, the MPEG-4 based on DCT could not transmit foreground and background in the different bit-rate simultaneously. However, based on the DWT of the 3D wavelet, our approach(Our-SVC) can achieve adaptive allocation of bit-rates in foreground and background, as shown in Fig.1.

This paper is organized as follows. Section II proposes an

978-1-4244-6516-3/10/$26.00 ©2010 IEEE

architecture of multi-region scalable coding system. Section III describes the key technologies of the system. The experimental results and analysis are given in Section IV. Finally, Section V is the conclusion.

(a) (b)

Fig. 1. The subjective quality comparison of MPEG-4 with Our-SVC at 112kbps, akiyo.(a) MPEG-4, (b) Our-SVC.

II. System Architecture

As shown in Fig.2, it describes the architecture of the proposed multi-region scalable video coding system. Its main modules are as follows:

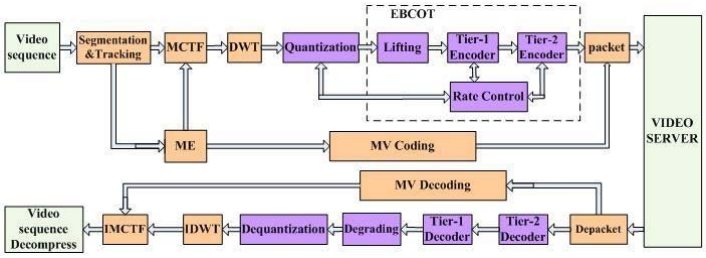

Fig. 2. The framework of Multi-region Scalable Video Coding System.

A. Region of motion tracking and segmentation module

Firstly, the user interactively draws a rectangular region to initialize the tracking region. Secondly, the paper uses particle filter tracking algorithm to track the multiple targets in a video, and finally gives multi-region boundaries and saliency weight respectively.

B. Regional motion estimation and MCTF

The main function of the module is to obtain regional motion estimation through several region of interest and a background regions in a GOP, and then to deal with the region of interest and the background pixel regions along the trajectory for time-domain wavelet filtering respectively, in order to eliminate the time-domain redundancy.

C. Region of the ROI adaptive wavelet bit-plane lifting module

After the 3D wavelet transform, it must find the corresponding coefficient of ROI in various frequency-domain sub-band. Bit-plane lifting is implemented according to the ROIs saliency weight adaptive scale these corresponding ROIs coefficient, the coefficients can have priority be encoding as the high weight region. In other words, according to the ROIs weight, the module could allow each of regions assigned to a different rate.

D. The adaptive rate control module (Tier1, Tier2)

Different bit-rates are allocated to the content of ROI under various conditions. The rate control module will integrate the current bit rate, video frame size to improve the factors, and optimize the bit allocation. In particular, it is through the higher ROI bit-plane lifting under a very low bit-rate, with more bit-rate allocation.

In the final packaging of the system, it would send the bit stream with a high degree of scalability. we can arbitrarily cut off the bit-plane coding bit-stream to realize the fine quality of scalable; hierarchical time-domain filtering process of time to realize scalable; while the space of wavelet transform to realize multi-scale decomposition of space is scalable. This flexible model will stream organizations to take full advantage of the current network bandwidth conditions, but also to meet the diversity of terminals, network heterogeneity, such as video communications and network needs.

III. Analysis of Key Technologies

A. Motion ROI

Some research has shown that photosensitivity is unevenly distributed in the retina. The human visual perception mechanism is similar to non-uniform sampling. Central fovea has a relavtively higher sensitivity. Human optic system will get more stimulation from this area. The visual attention objects in video sequences are mainly moving ones. With the help of this ingrained characteristics, it is possible to greatly enhance video transmission quality under limited bandwidth if we improve the quality of ROI during image or video coding process. Tracking of moving targets is a process of describing the state of objects and its changes over time. Two key questions are: how to represent the state, and how to represent change of the target state. Therefore, the video target tracking problem concentrates on how to express and locate objects, as well as how to design tracking filter.

Particle filter is generally known as a powerful tracking filter for its performance. The essence of particle filter is the use of Monte-Carlo method to simulate the process of recursive Bayesian estimation [4]. And the main idea is to use a set of weighted random sample of particles to express the probability distribution of state variables of the system, and to some guidelines for optimal use of the particle collection under the current system of calculating the estimated value of state variables. Katja et al. [5] would be the color distribution features and a combination of particle filters, the use of color measurement realize the characteristics of particle filter in the measurement process. Through the particle filter on the system state estimate, which can effectively adapt the complexity of the scene and has blocked the tracking environment in Katja's algorithm.

Katja's algorithm tracked the results of the presence of vibrating problems, such as ROI border region seriously affected

the uncertainty of the follow-up of the ROI region coding handle. The vibration is mainly due to: (1) measurement noise caused by vibration; (2) template is not very accurate, within the context of the template changes the impact. Also, because movement interested in the video has multiple moving targets. we need to realize a number of particle swarm tracking filter. To this end we focused on three conditions above for the particle filter algorithm of Katja's to do the following three improvements:

1) To achieve multi-target particle tracking, we use $S_{i,t}^{j}$ to describe the state of particle j in tracking filer i at time t, and then update the algorithm.

2) In the original algorithm introduced by the particle trajectory smoothing algorithm, through the records in a certain period and the particle motion trajectory and the value of these historical records need to be carried out smooth; In section III-B, it will be discussed in accordance with regional motion estimation methods, generally to a group of frames (GOP) within the state of smooth particles, and the corresponding template updates, such as Eq. (1) shown.

$$\overline{S}_{i,t}^{j} = \sum\nolimits_{n}^{GOP} S_{i,t}^{j(n)} \Big/ GOP \tag{1}$$

3) Using weighted mask. This approach corresponds to the unevenness of human vision. With regard of the calculation complexity, we update the weighted mask according to Eq.(2). The centered rectangle area α will be given a higher weight w_1, while surrounding area β will be given a lower weight w_2. The evaluation of weight is based upon the hypothesis that objects are most likely located in the center of rectangle area. Therefore, the weighted mask could effectively reduce the influence of background. In Eq. (2) , parameters are preset as follows: α= 60%, β= 40% ; $w_1 = 2$, $w_2 = 1$.

$$\text{w} = \begin{cases} w_1, & \text{a template centers outside the region}\alpha; \\ w_2, & \text{a template centers outside the region}\beta. \end{cases} \tag{2}$$

Based upon the improved Katja's particle tracking filtering, we achieved a robust and effective multi-target tracking algorithm. The implementation is illustrated in Fig.3. User will first initialize a rectangle area, which is labeled as region of interested, and then multi-target tracking algorithm will be adopted to obtain the object in the video sequence. We useRB_i to denote each ROI, and MR_i^l, to different resolution,$i \in [0, N]$,$l \in [0, L]$, where N is the number of objects, L is the number of scalable layer.

B. Region Motion Estimation and MCTF

Each frame of the original video is partitioned into N sectors, including multi-region and background. Background refers to non-ROI part, i.e.MR_0 .

According to the boundary description of the prospects and the background, RB_i will be the prospects for $(i = 1...N)$, and the background for $(i = 0)$ and the to carry out motion segmentation separately. Motion estimation methods are still using the traditional block-matching method such as the Eq.(3) shown. That is, for the prospect, as a separate video sequence to be carried out small, the coordinates of each pixel are used throughout the frame of absolute coordinates, and obtain motion vector tree of all the prospects. For background, it uses the entire frame to subtract the prospects for treatment, and then to get the background motion vector tree.

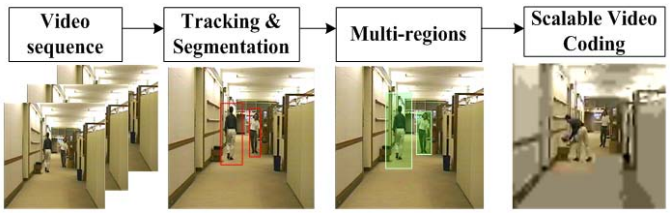

Fig. 3. Multi-region generated and SVC result diagram.

$$(dx, dy) = \arg\min \sum_{(x,y\in s)} |X_{cur}[x, y] - X_{ref}[x - dx, y - dy]|^p \tag{3}$$

Where $X_{cur}[x, y]$ is the current block pixel value; $X_{ref}[x - dx, y - dy]$ as the reference block pixel value; S is the search area; $|\bullet|^p$ is the matching criterion. When $p = 1$ when matching criterion is the sum of absolute difference (SAD), when $p = 2$, the matching criterion is the mean square error (MSE). After motion estimation, motion vector of block will be mapped to pixel, and motion compensation will be implemented. As for foreground, motion vector will be confined within region of interest. However, the background does not have such restrictions and motion vector will inevitably ingress foreground. In order to reconstruct the whole frame, we treat these pixels as irrelevant pixels. We use Le Gall 5-3 wavelet for filtering as Eq. (4)$\sim$ (5) :

$$\begin{aligned} H_i = {} & hweight[0]MAP_{2i+1\to 2i}(A_i) + hweight[1]B_i \\ & + hweight[2]MAP_{2i+1\to 2i+2}(A_{i+1}) \end{aligned} \tag{4}$$

$$\begin{aligned} L_i = {} & lweight[1]A_i + lweight[0]MAU_{2i\to 2i+1}H_{i-1} \\ & + lweight[2]MAU_{2i\to 2i-1}(H_{i+1}) \end{aligned} \tag{5}$$

where

H_i: High-frequency frame band after MCTF;

L_i: Low-frequency frame band after MCTF;

A_i: Pixel value in even frame;

B_i: Pixel value in odd frame;

MAP: Prediction steps, MA behind the brackets express the pixel along the trajectory and the corresponding pixels.

MAU: Update steps, MA behind the brackets express the pixel along the trajectory and the corresponding pixels.

$hweight[0] \sim [2]$: High-pass filter coefficients.

$lweight[0] \sim [2]$: Low-pass filter coefficients.

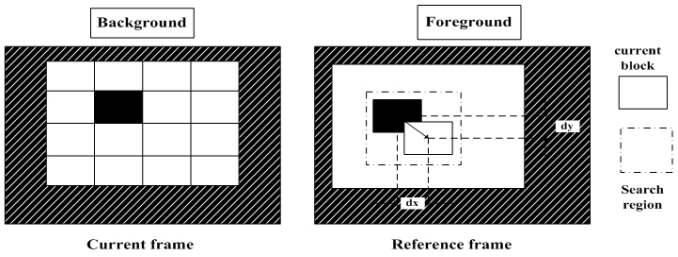

Fig. 4. Based on regional motion estimation.

C. Multi-region Wavelet Transform and Bit-plane lifting

After motion estimation, the motion vector tree is coded while both high and low frequency frame which are decomposed by spatial wavelets transformation. In order to acquire a comparatively higher ROI quality, information used to reconstruct the background needs to be decreased, which in turn means more bits will be allocated to ROI (MR_i^l, $i = 1...N$). We use Eq.(6) to generate multiple ROI mask. And then wavelets coefficient is lifted up according to Eq.(6).

In Eq.(7), M_b denotes current max bit-plane, and $numshift_i$ represents lifting bits for ROI. EBCOT is implemented after spatial wavelet transformation using bit-rate control, so as to allocate more bits to ROI and obtain better visual effects.

$$M(x,y) = \begin{cases} 1\ , & i = 1 \\ 2\ , & i = 2 \\ & ... \\ 0\ , & Background \end{cases} \tag{6}$$

$$numshift_1 \geq Max(M_b), \qquad i = 1...N \tag{7}$$

IV. Experimental Results and Analysis

In this paper, our experiment is based on "T+2D" structure that proposed in three-dimensional wavelet-based scalable video coding system [8]. Monitor on standard tests of CIF video format (resolution: 352×288), in improvement Katja [5] color characteristics of particle filter algorithm based on the generation of multi-region and test in the interest scalable video coding system.

Fig.5 gives the bit-rate under a variety of multi-region and non-ROI coding ROI in the same region of the contrast curve of PSNR values. The experimental results for the relevant sample are shown in Fig.6. From the experimental results we can draw the following conclusions:

1) From Fig.5 comparison of the PSNR curve, we find that multi-region scalable video coding in the ROI region PSNR values larger than non-ROI method PSNR value about 4dB. It shows that the performance of our approach is effective, in particularly the situations of lower bit rate.

2) From Fig.5 and Fig.6(b) we find that our multi-scalable coding ROI method is more suitable for low bit rate to improve subjective and objective quality of video transmission. As in Fig.5 as seen from the PSNR values when 1024Kbps see more

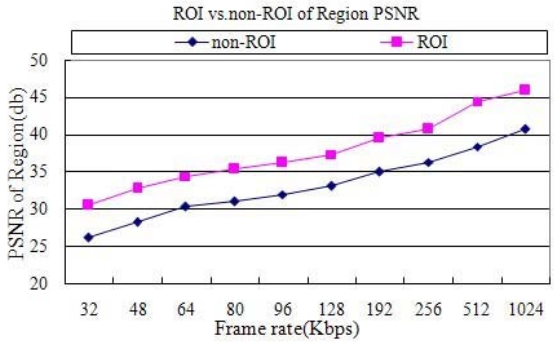

Fig. 5. The PSNR comparison of Multi-ROIs with non-ROI coding.

Fig. 6. Multi-region Scalable video coding results of monitor sequence.(a)monitor sequence, Multi-target tracking results with PF method;(b) 1024Kbps, Multi-region coding; (c) 48Kbps, non-ROI coding; (d) 48Kbps, Multi-region coding;(e) 128Kbps, Multi-region coding; (f) 256Kbps, Multi-region coding.

ROI coding circumstances PSNR higher than non-ROI coding situation 5.1dB, but in Fig.6(b) differences in the subjective quality is not obviously difference.

3) Under a lower rate 48Kbps, see Fig.6(c) in the context of the same area code and more interested in blocking, video subjective quality low in the ROI; While in Figure.6(d) the contents of the multiple region of interested are still be

expressed clearly. While at the same time, the region is also the expense of background details for the price.

4) Fig.6(e), (f) showed that in the 128Kbps and 256Kbps circumstances respectively, both are to ensure the objective and subjective quality of multi-region of interest while being able to gradually improve the quality of the background of the regional results. The integrated experimental results show that the proposed multi-region scalable video coding method, is especially feasible and effective in the case of low bit-rate.

V. Conclusion

This paper describes a multi-region scalable video coding method, by analyzing the motion video and tracking based on the given multiple moving targets as the region of interest video content, then through using regional MCTF algorithm for motion estimation as well as regional wavelet transform and bit-plane lifting scheme, and finally Tier1, Tier2, etc. The model has the abilities of real-time and robustness, and is easy to realize, can support CIF, QCIF, 4CIF etc. This model can also be rate limited, especially in low bit rate cases, the priority transmission of interested regional video content with high-quality, thereby reducing as much as possible is appropriate in the mobile video and surveillance video applications and so on because the network bandwidth can not be brought about by the impact of video. However, in the case of existing occlusion motion object, or video motion background is complex, and the model has limitations. Therefore, also in real-time multi-target tracking algorithm, motion estimation of regional areas is necessary to be further studied for the purpose of improving the content quality of scalable video.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant Nos. 60635050 and 60875008, the National Basic Research Program of China (973 Program) under Grant No. 2010CB327902. At the same time, the authors thank the anonymous reviewers for their perceptive comments.

References

[1] T. M. Bae, T.C. Thang, D. Y. Kim, Y. M. Ro and J. G. Kim, "Spatial scalability of multiple ROIs in scalable video coding," *Proceedings of the SPIE*, vol. 6074, pp.53-60, 2006.

[2] Y. Kim, S.H. Jin, T.M. Bae and Y.M. Ro, "A selective video encryption for the region of interest in scalable video coding," *IEEE TENCON 2007- IEEE Region 10 Conference*, pp.1-4, 2007.

[3] Q. Chen, G. T. Zhai, X. K. Yang and W.J.Zhang, "Application of scalable visual sensitivity profile in image and video coding," *IEEE International Symposium on Circuits and Systems*, pp.268-271, 2008.

[4] J. R. Xue, N. N. Zheng, J. Geng and X. P. Zhong, "Tracking multiple visual targets via particle-based belief propagation," *IEEE Transactions on System, Man, and Cybernetics Part B*, vol. 38, no. 1, pp.196-209, 2008.

[5] K. Nummiaro, E.K. Meier and L.J.V. Gool, "Object tracking with an adaptive color-based particle filter," *IEEE International Symposium for Pattern Recognition of the DAGM*, pp.353-360, 2002.

[6] J. R. Ohm, "Advances in scalable video coding," *Proceedings of the IEEE*, Vol. 93, no. 1, pp.42-56, 2005.

[7] Y. J. Wu and Woods, J.W., "Scalable motion vector coding based on CABAC for MC-EZBC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 6, pp.790-795, 2007.

[8] R. Xiong, X. Ji, J. Xu, and F. Wu, "MSRA scheme for SVC CE1," *ISO/IEC JTC1/SC29/ WG11 MPEG 70th meeting, M11320*, Palma, Oct. 2004.

[9] "MPEG-4 requirements version 4.0," *ISO/IEC JTC1/SC29/WG11, MPEG-4 Requirements Group Stockholm*, Sweden, 1997.