Jr. of Industrial Pollution Control 33(2)(2017) pp 1555-1559
www.icontrolpollution.com
Research Article

MOVING OBJECT AWARE VIDEO COMPRESSION

PHU TRAN TIN [1], ANH VU LE [2*]

[1]Faculty of Electronics Technology, Industrial University of Ho Chi Minh City, 12 Nguyen Van Bao Street, Ho Chi Minh City, Viet Nam

[2]Optoelectronics Research Group, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

(Received 07 August, 2017; accepted 20 September, 2017)

Key words: Moving object, Video compression resizing, PSNR

ABSTRACT

Moving object is detected by simple spatial and temporal gradients in GOP (Group of Frame) of H.264/AVC such that the MOOI (Moving-Object-of-Interest) is remained as much as possible while the non-MOOI is squeezed. This resizing process to original video is applied as a pre-processing for standard video compression. Specifically, more image data of the original video will be dropped by our non-uniform subsampling as the distance from the center of the MOOI increases. Experimental results show that our MOOI aware video compression can preserve the original visual quality of the MOOI even for low bit-rate applications.

INTRODUCTION

Video compression plays an important role in digital video surveillance and DVR (digital video recording) systems (Venkatraman and Maku, 2009; Kim, *et al*., 2012). A major concern issues of the above applications is how to efficiently compress the long hours of video data with high compression ratios. That is, the problem is how to maintain high compression ratios while preserving visual quality of the important objects in the video. The senses include the static and moving objects regions. Obviously, moving objects are considered as important regions in surveillance videos. Therefore, the desired requirements are that one can separate the important MOOI (Moving-Object-of-Interest) from the unimportant non-MOOI to treat them separately for compressions. To this end, object segmentation and tracking processes can be applied (Spagnolo, *et al*., 2006; Kim and Hwang, 2002; Chien, *et al*., 2002). However, these methods need to identify accurate object boundary, which in turn requires computationally expensive segmentation and tracking processes.

Our approach for the above problem is to roughly identify the center of the MOOI and the bounding box surrounding the MOOI for each GoF (Group-of-Frame) with simple spatial and temporal gradient energies, where the bounding box of the detected MOOI is fixed for all frames in the GoF. Here, the MOOI detection is based on the representative frame of GoF in (Nguyen and Won, 2013) and it is the area of temporal and spatial saliency. Now, the areas of the MOOI and non-MOOI are identified and we apply a logarithmic nonlinear transformation of (Won and Shirani, 2011) at the center of the MOOI combining with the linear resizing of (Le, *et al*., 2014) to reduce the size of the image frame such that the image data within the MOOI is intact while those in the non-MOOI are reduced as much as possible. Then, the size-reduced image frames undergo the standard H.264/AVC compression for further compressions. At the receiver, the compressed bit-stream is first applied to H.264/AVC decompression to reconstruct the size-reduced video and then the invertible logarithmic transformation and an interpolation scheme are applied to restore the video data with the original image size. Note that image pruning scheme with image down-sampling as a

***Corresponding authors email:** leanhvu@tdt.edu.vn**

preprocessing step of video compressions has been also proposed in (Vo, *et* al., 2010), where one of the two consecutive image lines (i.e., even or odd lines) is to be dropped for image size reduction. Since the line dropping is limited for one of two consecutive lines and the criterion for line dropping is based on the LMSE (Least Mean Square Errors) of the interpolated image data, it is hard to treat the MOOI and the non-MOOI separately.

DETECT MOVING OBJECT OF INTEREST AND VIDEO COMPRESSION APPROACH

Detecting moving object by motion vector is the burden process. To simplify, MOOI is detected by using spatial and temporal gradients in each GoF of H.264 to prevent artifact in temporal domain. A spatial gradient map $S_g^t(i,j)$ of even frame I^t size N_r x N_c of a GOF k^{th} including N_0 frames from frame m_0 (suppose is even number) is a sum of spatial gradients within a window 2Ψ+1 as (1)

$$S_g^t(i,j) = \sum_{i=-\psi}^{\psi}\sum_{j=-\psi}^{\psi} \left|\partial I^t / \partial i\right| + \left|\partial I^t / \partial j\right| \quad (1)$$

We can compute the temporal saliency cost $S_t^n(i,j)$ in (2) as temporal gradient changes between spatial gradient maps of two executive even frames:

$$S_{tsc}^t(i,j) = \left| S_g^t(i,j) - S_g^{t+2}(i,j) \right| \quad (2)$$

Different from [6] which proposed gray level representative frame, in our paper a temporal gradient representative frame TRF_{GoF}^k of all even frames of a k^{th} GoF is defined as (3)

$$TRF_{GoF}^k(i,j) = \left\{ \begin{array}{l} S_{tsc}^{t*}(i,j) \,|\, t* = \underset{t=m_o:2:N_0-2}{\arg\max} \left\{ S_{tsc}^t(i,j) \right\}, \\ 1 \le i \le N_r - 2\psi - 1, 1 \le j \le N_c - 2\psi - 1 \end{array} \right\} \quad (3)$$

Because motion pixel (i, j) will present higher $TRF_{GoF}^k(i,j)$ value than non-motion pixel, the MOOI of our proposed method is the window $(2w_r+1)$ $(2w_c+1)$ where sum of $TRF_{GoF}^k(i,j)$ within this window yields the maximum value and center $C(C_r, C_c)$ of MOOI is the center of this window. (Fig. 1) shows the result for one moving object detection and two moving object detection. The method proposed in (Le, *et al.*, 2014) are used determine the optimal windows size and center. By using this method, the size of window which encapsulates the moving objects can be adjust to fit the much moving areas in GOF (Fig. 1).

After MOOI is defined, we apply the non-uniform logarithmic transform as (Won and Shirani, 2011) to reduce original size N_r x N_c to size M_r x M_c at encoder at a predefined reduced rate. The (4) is used to reduce row size $N_r \rightarrow M_r$ then the same way to column size.

$$S_c[n_r] = \left\{ \begin{array}{ll} C_r' - \beta_r \ln(\alpha_r d(n_r) + 1, & if\ n_r < C_r \\ C_r' + \beta_r \ln(\alpha_r d(n_r) + 1, & if\ n_r > C_r \end{array} \right. \quad (4)$$

$$where\ C_r' = round[\beta_r \ln(\alpha_r C_r + 1)], d(n_r) = |C_r - n_r|,$$

$$and\ \beta_r = (M_r - 1) / \ln(\alpha_r C_r + 1)(\alpha_r (N_r - 1 - C_r) + 1),$$

C_r C'_r are row index of center of MOOI of original frame and reduced frame respectively. $d(n_r)$ is the absolute distance from C_r to pixel position n_r in original frame. β converts the original size N_r to reduced size M_r. α_r is responsible for the degree of expansion MOOI after row reduction. Based on the

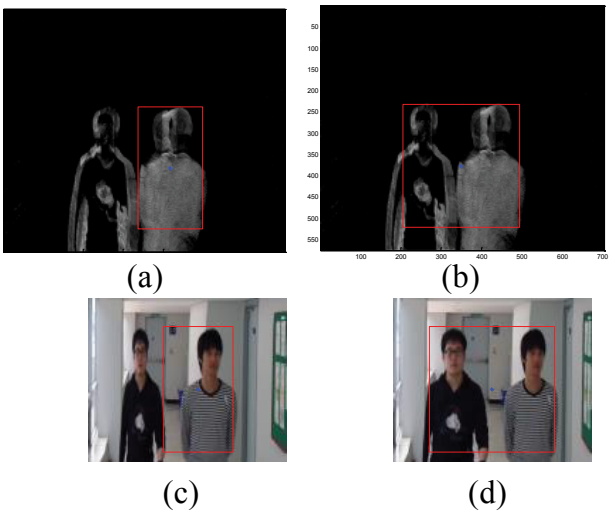

Fig. 1 Optimal window size and reduced frame with intact MOOI. (a) optimal window for one moving object (b) optimal window for two moving objects (c) reduced frame for one moving objects (d) reduced frame for 2 moving objects. (a) Detected one moving object, (b) Detected two moving objects, (c) Frame reduction keeping one moving object, (d) Keeping 2 moving objects.

characteristic of logarithmic transform equation 4, the nearer the pixel positions to center of MOOI in original video frame, the more density of pixels in down-sampled frame grid (Fig. 2).

In this paper, a method is proposed to determine row α_r to make almost no reduction or expanding within MOOI after row reduction, which is illustrated as (Fig. 2). The following (5) is the criterion to determine w_r. The same way can be used to find column α_c for w_c.

$$| d'(n_r) - d(m_r) | = | \beta_r \ln(\alpha_r w_r + 1) - w_r | \le 1 \quad (5)$$

Recall: $\beta_r = (M_r - 1) / \ln(\alpha_r C_r + 1)(\alpha_r (N_r - 1 - C_r) + 1),$

$$(5) \Leftrightarrow \left| \frac{(M_r - 1)\ln(\alpha_r w_r + 1)}{\ln((\alpha_r C_r + 1)(\alpha_r (N_r - 1 - C_r) + 1))} - w_r \right| \le 1$$

$$\left| \frac{(M_r - 1)\ln(\alpha_r w_r + 1)}{\ln(\alpha_r C_r + 1) + \ln(\alpha_r (N_r - 1 - C_r) + 1)} - w_r \right| \le 1$$

The condition in (6) is the results after appling Taylor approximation.

$$\left| \frac{(M_r - 1)(\alpha_r w_r - \frac{\alpha_r^2 w_r^2}{2})}{(\alpha_r C_r - \frac{\alpha_r^2 C_r^2}{2}) + (\alpha_r (N_r - 1 - C_r) - \frac{\alpha_r^2 (N_r - 1 - C_r)^2}{2})} - w_r \right| \le 1 \quad (6)$$

$$where: 0 < \alpha_r w_r \le 1, 0 < \alpha_r C_r \le 1, 0 < \alpha_r (N_r - 1 - C_r) \le 1,$$

After doing mathematical expression the α will be calculated as (7)

$$\frac{2((w_r - 1)(N_r - 1) - (M_r - 1)w_r)}{(C_r^2 + (N_r - 1 - C_r)^2)(w_r - 1) - (M_r - 1)w_r^2} \le \alpha_r \le \frac{2((w_r + 1)(N_r - 1) - (M_r - 1)w_r)}{(C_r^2 + (N_r - 1 - C_r)^2)(w_r + 1) - (M_r - 1)w_r^2} \quad (7)$$

To refine estimated row alpha more accurately, an iterative method can be used with an above selected row alpha pays a role as initial row alpha. Iterative step=0.001 and Iterative times = interactive max It_{max}, interactive loop will stop if condition of (5) is satisfied (Fig. 3).

Finding optimal alpha by above method can assure that the pixels in reduced frame at the boundary of predefined window is kept as same as original

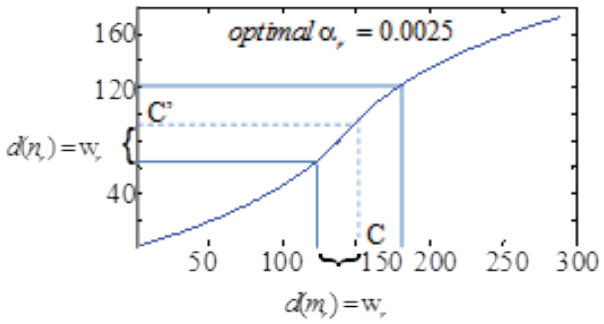

Fig. 2 Size reduction on one direction with no change within MOOI w=32.

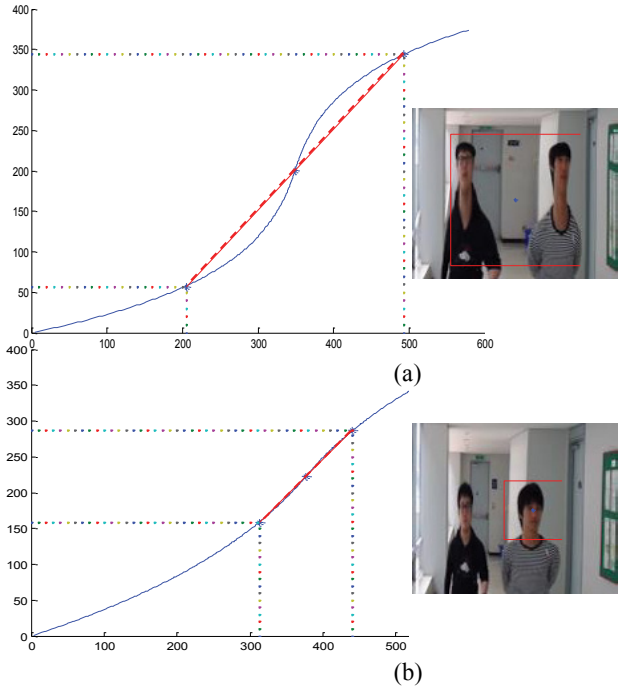

Fig. 3 Distortion within big window using logarithmic transform, blue line is logarithmic transform and red line is linear transform. (a) Big window (b) Small window.

pixels. However, in case of big windows, logarithmic transform as (7) makes the pixels within this window is distorted especially if window size is too big (Fig. 3). To remove this artifact, instead of logarithmic transform, we propose a linear transform (the red line) will be used for all pixels within MOOI window if the windows is larger than predefined threshold W_m. To this end, for the big window, our transform is the combination between logarithmic transform within non-MOOI and linear transform within MOOI (Fig. 4).

After size reducing, conventional H.264 is used to compress reduced video at a predefined bit rate. Note that reduced frames have MOOI almost same as original frames and these size is less than original frames. This way makes ROI can be compressed by lower compression ratio at the same bit rate than original H.264. The main advantage of logarithmic transform (Won and Shirani, 2011) is that it is invertible ability to retrieve the original size after reducing. This means, at the decoder for non-MOOI region, expanding reduced size to original size is done firstly for horizontal then vertical size by an invert mapping (8). $d'(m_r)$ is the absolute distance from C'_r to pixel position m_r. If m_r in equation (8) is non-integer in the frame grid, one interpolation operator likes bilinear can be used to find the nearest integer pixel value. Considering MOOI region, because the parameters

center of MOOI and MOOI size are known as well as MOOI in reduced frame is almost same as original frame, pixels are matched one to one from reduced frame to expanded frame. Note that, the parameters $\alpha_r, \alpha_c, w_r, w_c, C(C_r, C_c)$ and reduction rate are sent to decoder as side information to expand reduced frame to original size.

$$S_{C'}^{-1}[m_r] = \begin{cases} C_r - \dfrac{\exp(\frac{d'(m_r)}{\beta_r}) - 1}{\alpha_r}, & if\ m_r < C_r' \\ C_r + \dfrac{\exp(\frac{d'(m_r)}{\beta_r}) - 1}{\alpha_r}, & if\ m_r > C_r' \end{cases} \quad (8)$$

$$where: d'(m_r) = |C_r' - m_r| \ and\ C_r = round[\exp(C_r'/\beta_r - 1/\alpha_r].$$

EXPERIMENT AND RESULTS

The surveillance video sequences from the (http://ivylab kaist.ac.kr/demo/vs/dataset.htm) were used. In our experiment Wm=100. At the encoder, frames in each GOF defined by H.264 parameter are down-sampled by our method in section II and LMSE method in paper (Vo, *et al.*, 2010) then compressed at the fix bit rate. At the decoder, they are up-sampled by bilinear interpolate. Frame reduction ratio 30% and bitrates from 50-200 kbps are tested to find the critical bitrate. Proposed method is compared with conventional H.264. (Fig. 4) gives the results of difference kinds of transforms. Our proposed method (the combination of logarimic transform and linear transform) can preserve the moving object intact. Visualize results of MOOI are provided in (Fig. 5) for one frame of hallway sequence. As one can see, proposed method outperforms than H.264 and LMSE in MOOI in terms of PSNR and visualize. Look at the person face that is detected as MOOI of 450th frame of Stair sequence, as the result of keeping MOOI

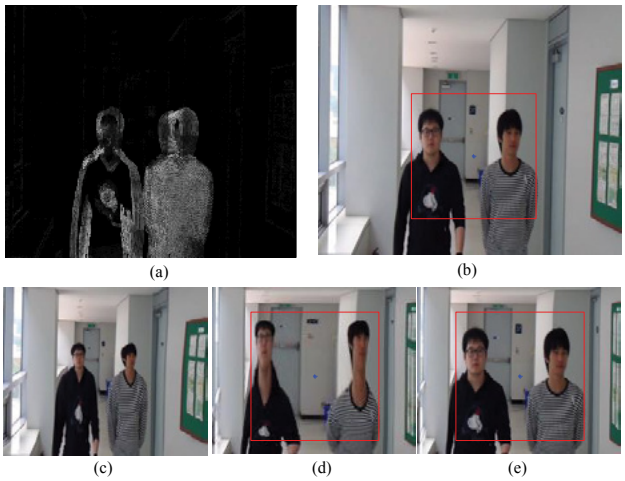

Fig. 4 Reduction 30%. (a) Representative temporal gradient frame (b) Original (c) LMSE (d) Reduce logarithmic in window (e) Reduce linear in window.

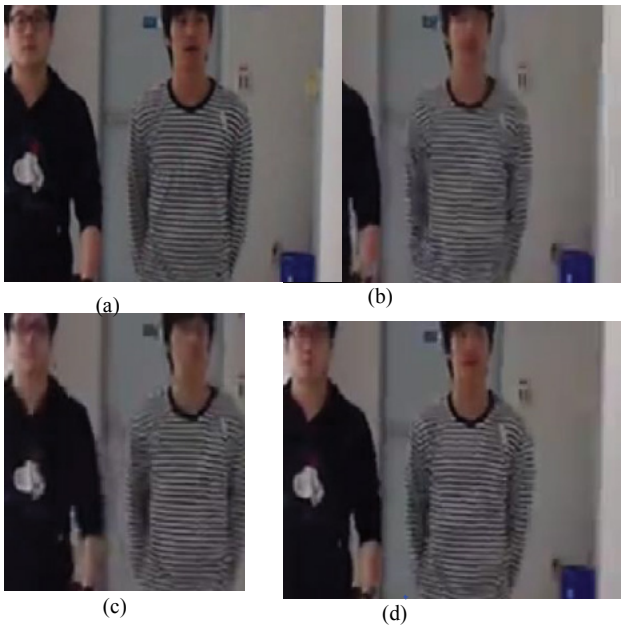

Fig. 5 Zoom in expanded MOOI of 450th frame of Hallway sequence: 30% reduced frame bitrate 100 kbps. (a) Original (b) H.264 (c) LMSE (d) Proposed method.

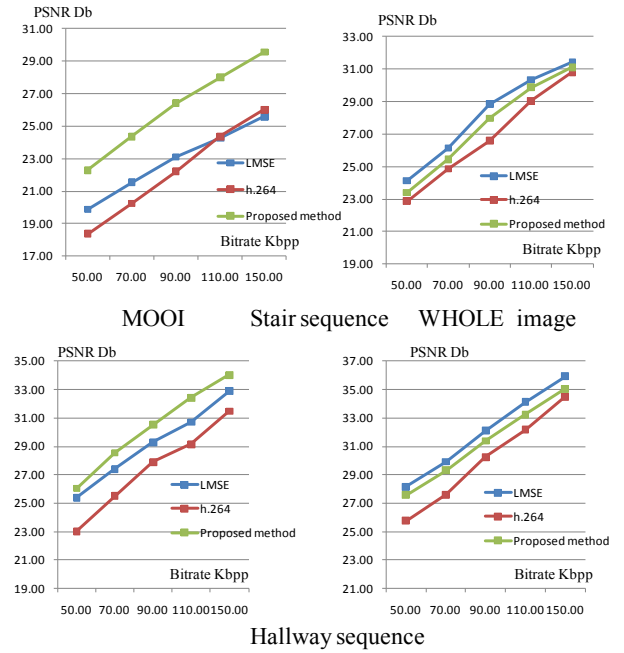

Fig. 6 Rate and distortion R&D for MOOI regions and whole image.

intact after frame reduction step and using lower compression ratio at same bitrate, proposed method can be recognized more clearly despite very high compression ratio. In other hand, His face cannot be seen clearly if LMSE or conventional H.264 are used. LMSE is even worse than H.264 because it take care only complexity areas. (Fig. 6) provides the rate and distortion curves. Although our method yield lower PSNR than LMSE for whole frame, our method is the

best method at critical bitrate lower than 150kbps in MOOI region (Fig. 5 and 6).

CONCLUSION

The contributions of this paper are the introductions of roughly detecting moving object of interest (MOOI) method and keeping decoded MOOI clearly in spite of high compression ratio. A preprocessing step that combines logarithmic transform and linear transform to reduce video frames resolution is proposed. The parameters alpha of logarithmic transform is optimal so that MOOI is kept intact after reducing step. Experimental results show the decoded videos yield the best results in terms of PSNR for MOOI about 3dB higher than LMSE and visualize comparison. The proposed approach is the good idea to applications where video steam should compression at high compression ration like surveillance scenes, video calls.

REFERENCES

Chien, S.Y., Ma, S.Y. and Chen. G.L. (2002). Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*. 12(7) : 577-586.

http://ivylab kaist.ac.kr/demo/vs/dataset.html.

Kim, S., Lee, B.J., Jeong, J.W. and Lee, M.J. (2012). Multi-object tracking coprocessor for multi-channel embedded DVR systems. *IEEE Transactions on Consumer Electronics*. 58 : 1366-1374.

Kim, C. and Hwang, J.N. (2002). Fast and automatic video object segmentation and tracking for content-based application. *IEEE Transactions on Circuits and Systems for Video Technology*. 12(2). 122-129.

Le, A.V., Jung, S.W. and Won, C.S. (2014). Non-uniform video size reduction for moving objects. *The Scientific World Journal*.

Nguyen, H.T. and Won, C.S. (2013). Video retargeting based on group of frames. *Journal of Electronic Imaging*. 22(2) : 023023-023023.

Spagnolo, P., Leo, M. and Distante, A. (2006). Moving object segmentation by background subtraction and temporal analysis. *Image and Vision Computing*. 24(5) : 411-423.

Venkatraman, D. and Makur, A. (2009). A compressive sensing approach to object-based surveillance video coding. IEEE International Conference on Acoustics, Speech and Signal Processing.

Vo, T., Sole, J., Yin, P., Gomilaan, C. and Nguyen, Q. (2010). Selective data pruning-based compression using high-order edge-directed interpolation. *IEEE Trans. Image Process*. 19(2) : 399-409.

Won, C.S. and Shirani, S. (2011). Size-controllable region-of-interest in scalable image representation. *IEEE Trans. Image Process*. 20(5) : 1273-1280.