Background Separation Encoding for Surveillance Purpose by using Stable Foreground Separation

Tomohiro Kitagawa, Tomoya Koseki and Takao Nishitani
Tokyo Metropolitan University, Hino 191-0065 Tokyo Japan
E-mail: {kitagawa@sd., koseki@sd., nishitani@}tmu.ac.jp

***Abstract*— An efficient encoding approach by separating background/foreground videos from a single video is proposed for surveillance applications. The required bit-rate becomes less than a half of that by a standard H.264 approach. The reduction comes from the employment of the low resolution background pictures with a low frame-rate. In addition, the foreground video is automatically handled so that the information amount becomes almost zero during the periods of no foreground objects. Therefore, storing videos or collecting multi-point surveillance videos at a surveillance center will save storage capacity or communications costs.**

I. INTRODUCTION

Although the international standard of MPEG-4 core profile [1] appeared long time ago as an object based coding, no such services become active, due to the difficulties of object extraction from video contents. However, in outdoor surveillance, an object based coding is preferable for observing precise foreground objects under reduced information to send or to store. Several attempts have been reported for background separation approaches [2][3], but the most of such video coding approaches use the background as a still picture. In addition, the camera resolution used there is set to the small SIF/CIF format and encoding bit-rate is around 64kbps or low. An example of high resolution video using background separation can be seen in MPEG-4 core profile, where a scenery around a camera is sent only once at the beginning of encoding as a background. Again, it is a still picture, used in the sprite approach. The actual background scenery is replaced by a part of the still background. Background update functions have not been specified, partly because every still picture requires a lot of information, covering the wider background scenery.

Nowadays, HDTV cameras become cheap and broadband networks become available. HDTV surveillance becomes attractive for precise object observation in out-door surveillance, but a lot of motion in a background, such as foliage of trees blowing wind and reflections from rivers and windows, can be easily observed. Such a background is called a dynamic background and generates a lot of information. The conventional approaches might segregate such a dynamic background into foreground with a curious shape, and the actual foreground quality might become worse.

This is a trial of an object based coding for high resolution surveillance applications, where background information does not affect foreground object quality and drastic reduction of required information can be achieved, when no foreground object appears. The approach employs a couple of H.264 encoders with the Variable Bit Rate (VBR) mode and a recently developed spatio-temporal GMM (Gaussian Mixture Model of backgrounds) approach for background separation [4]. Two encoders are required, because the background is also considered to be a video with motion. The background information of weathers on sunny, cloudy, raining or snowing, and of time periods of dawn, day time, evening or night, is also considered to be important surveillance items, but it should be heavily compressed. Foreground segmentation based on the spatio-temporal GMM gives stable segmentation under dynamic backgrounds by introducing texture statistics as well as temporal statistics of backgrounds.

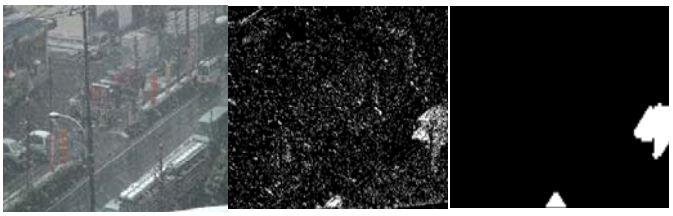
The approach first decomposes an HDTV video into a background video and a foreground video. The frame-rate and the frame resolution of the background video is reduced for lowering bit-rate from the background. The foreground video is directly fed into another encoder. The time period of no foreground automatically produces almost negligible bit-rate, due to the intra-prediction and the VBR mode in H.264.

The paper is organized as follows. In Section 2, the spatio-temporal GMM foreground segmentation is briefly reviewed. Section 3 describes our low bit-rate coding approach and in Section 4 some experimental results are shown.

II. SPATIO-TEMPORAL GAUSSIAN MIXTURE MODEL

A Gaussian mixture model (GMM) is widely used for the foreground separation from a dynamic background video [4], but it models a temporal probability function of every input pixel value. Due to the pixel base processing, it results in a heavy processing amount. In addition, due to the independent pixel processing without considering neighbor pixel status, the stability is not so high, especially in global light changes. In contrast, the spatio-temporal GMM (STGMM) used here has the capability of realizing a stable foreground separation, with a reduced operation amount in the following way.

1. An input frame is first divided into small blocks and these blocks are transformed into spectrum domain. A set of multiple block sizes, from 4x4, 8x8, 16x16, 32x32 to 64x64, is employed. The employment of multi-resolution processing makes the segmentation stable.
2. GMM is established in every block by making low band



(a) Original (b) GMM (c) STGMM

Fig. 1. Foreground Separation Results. Only running cars are segmented by the STGMM in (c).

spectrums. Only two parameters from every block are employed: the DC parameter and a vector parameter of horizontal, vertical and diagonal low band components in every block. As only two parameters are employed in every block, the processing amount becomes drastically reduced to 10% of the original one.

3. The foreground decision is carried out in every block size. The final decision is carried out by combining decisions results from every block size. As a result, the foreground detection area is composed of a combination of several block size decisions. The minimum block size of 4x4 fits well in the H.264 coding.

Thanks to the spatial information from spectrum components and to the multi-resolution processing, STGMM can stably extract the foreground objects under noisy background. Fig.1 shows an example of the effects, where a car is running in the heavy snow. The per-pixel GMM detects all the snow falls, but the STGMM can detect only the running car in Figs. 1 (b) and (c), respectively.

III. ENCODING APPROACH

A HDTV video is first decomposed into a background video and a foreground video by the stable STGMM separator, shown in Fig.2. The resultant background video is subject to be reduced both in the spatial resolution and in temporal resolution, even if some dynamic backgrounds are included. This processing contributes to reduce information amount from the backgrounds. On the other hand, the foreground video is produced by putting foreground objects onto a mono-tone background. These two videos are encoded by using a couple of H.264 encoders in the VBR mode. Then, the foreground video produces only a small amount of information, when a foreground object is not detected. This is because intra-frame and inter-frame predictions in H.264 work well over all frames during mono-tone background video periods. Although the decoded background video lacks both spatial and temporal resolution, the environmental changes of blowing strong winds or starting heavy rain can be still observed by this approach. In the following, more precise processing in background/foreground videos is described.

A. Background Sequence Generation

In order to reduce the required bit-rate on a background

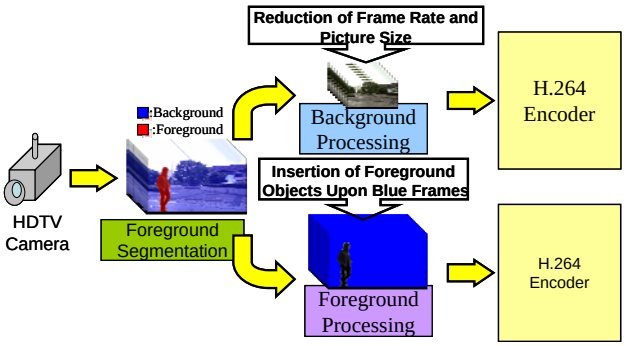
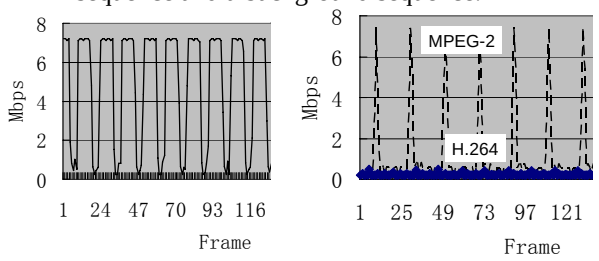


Fig. 2. Background Separation Coding.

A video sequence is decomposed of a foreground sequence and a background sequence.



(a) Still Picture Frames (b) Blue frames

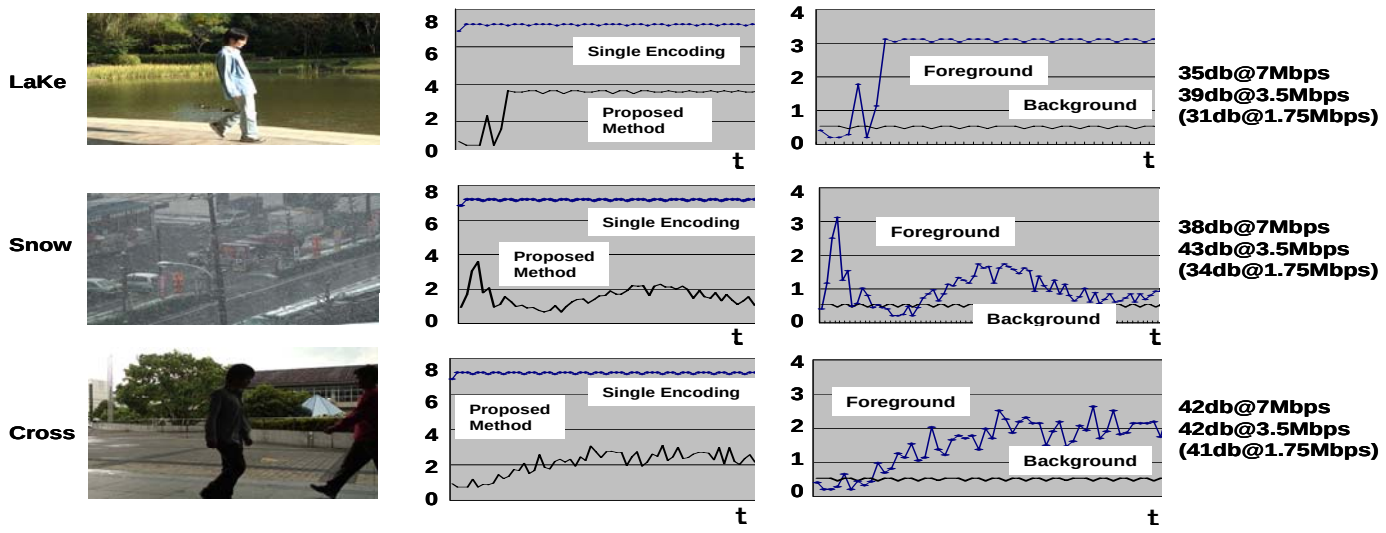
Fig. 3. Bit Rate Variations.

video, the background video is reduced to have a smaller frame size with slower frame-rate. As the input video frame has 1920x1080 pixels, the reduced frame size is set to 960x540 for example: 1/2 reduction in both vertical and horizontal directions, but this frame size is still larger than the conventional NTSC frame of 720x480 pixels. As the background is not a major part to be observed, the employed frame-rate is set to 7.5 fps, 1/4 frame rate reduction of a conventional NTSC video. Total reduction reaches 1/16. The background video is fed into a H.264 motion video encoder for dynamic background purposes. The reduction of the frame size and the frame rate on the background video contributes lowering the bit-rate for encoding, because H.264 encoders generate rather high bit-rates at every I-picture (Intra picture),. Fig.3 (a) shows an example of still picture frames. The instantaneous bit-rate on still pictures reaches the allowable maximum bit-rate at every I-picture even in the VBR mode.

Additional processing is required in the background video for fulfilling the holes, generated by the extraction of foreground objects. This is because four foreground frames appear during one background frame period, due to the difference of frame rates between foreground and background videos. Data to bury the hole are brought from the latest background areas, corresponding to the locations of holes.

B. Foreground Sequence Generation

The foreground video is composed of mono-tone background frames and the extracted foreground objects. The employment of a mono-tone background is partly because of the control key for the final output video, synthesized from the foreground and background videos (normally called



(a) Example Picture (b)Standard/Proposed Encoding Bit Rate (c) Foreground/background Bit Rate (d)Foreground PSNR

Fig. 4 Evaluation of bit rates on a 7Mbps standard H.264 encoder and a proposed encoder operated at 3.5 Mbps by using three typical HDTV videos of Lake, Snow and Cross. Foreground PSNR comparison includes 1.5Mbps proposed approach.

chroma-key synthesis) at the merging process of two H.264 decoded videos, and is partly because of efficient compression by a H.264 encoder during no foreground objects. A sequence of blue pictures generates only 0.2 Mbps in the Transport Stream (TS), due to the high prediction capability of intra-/inter-frame prediction on consecutive blue pictures in H.264. Fig. 3(b) shows the bit-rate variations of such a sequence, encoded by MPEG-2 and H.264, where the MPEG-2 performance is depicted by a dot line and that of H.264 is the bold line near the x-axis. The MPEG-2 line periodically shows high peaks, corresponding to the I-picture positions, but very little peaks can be seen in the H.264 line. As a result, during no foreground object periods, the foreground sequence generates almost negligible bit-rate of 0.2 Mbps.

IV. Experimental Results

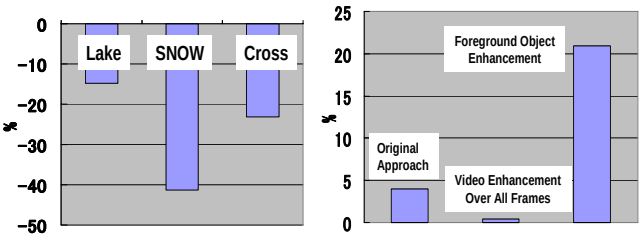
The experiments were carried out to evaluate a background separation coding approach, compared to a standard H.264 coding under the VBR mode. The standard H.264 is set to 7 Mbps of both maximum and average bit-rates. This is almost the same to the CBR (Constant Bit Rate) mode. However, no-stuffing bits are employed, even if the generated coding bit amount does not reach the average. For the proposed approach, the foreground video encoder is set to the VBR mode of 3 Mbps and 2.5 Mbps for the maximum and average bit-rates, respectively, while the background encoder is set to 0.5 Mbps for both the maximum and average bit-rates. Therefore, the proposed approach generates the total bit-rate of 3.5 Mbps at most, which is half a bit-rate of the standard one. As the background information before coding becomes 1/16, 0.5 Mbps for a background video is considered to be enough. Three HDTV test sequences, all of which have dynamic backgrounds shown in Fig. 4, are encoded by using a Main Concept software H.264 encoder.

The second column in Fig. 4 shows the difference between the 7 Mbps standard H.264 and the proposed approach, where the graphs of the proposed one are depicted by the total bit-rate of foreground and background information. The standard H.264 always shows 7 Mbps in every frame, but the proposed approach shows around 3.5 Mbps at peak positions.

The third column in Fig. 4 shows bit-rate difference between background and foreground videos in the proposed coding. As every video includes a dynamic background, the background videos always show the maximum bit-rate of 0.5 Mbps in every video. On the contrary, foreground video bit-rate varies less than 3 Mbps, depending on the conditions on foreground object such as sizes and object darkness. As the background in the foreground video is replaced with the blue scenery, this background area generates only negligible information, shown in Fig. 3(b).

The fourth column in Fig. 4 shows foreground object PSNR (Peak SNR) on the foreground objects, encoded by the 7 Mbps H.264 and that of the proposed approach at the highest bit rate of 3.5 Mbps. Although the proposed approach requires only 50% of the H.264 bit-rate, the PSNR of the objects in the proposed approach shows nearly 5 dB higher performance, except the Cross. The objects in the Cross video are dark and, therefore, the object PSNR has been saturated. The object PSNR on further bit rate reduction to 1.75 Mbps encoding based on the proposed approach has been shown in the parenthesis, where the foreground video is encoded at 1.5 Mbps and the background video is 0.25 Mbps with further frame rate reduction of 3.75 fps. Only 5 dB degradation has been observed, comparing with the 7 Mbps H.264.

In order to clarify the total picture quality, the subjective evaluation between the 7 Mbps H.264 and the 3.5 Mbps proposed approach were carried out by using the DSCQS



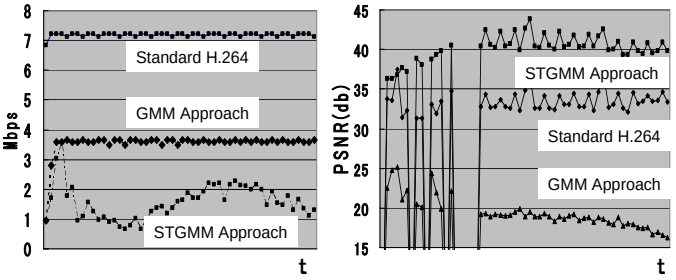
(a) Three videos get worse Scores than H.264. (b) Foreground Enhancement becomes No.1.

Fig. 5. Subjective test results on the proposed approach. The reference quality is set to7 Mbps H.264.

(Double Stimulus Continuous Quality Scale) method. Fig. 5(a) shows the test results of three videos in Fig. 4. As the bit-rate used for the background video is limited to only 0.5 Mbps and the total bit-rate in the proposed approach is a half or less, the subjective quality of the background separation approach becomes worse than that of the H.264. The scores of three videos clearly show that every picture quality of the proposed approach is less than that of the H.264. Especially, the snow sequence is worst, due to the heavy compression on the dynamic background of many snow falls. However, clear objects are still there and blocky distortions cannot be observed in the backgrounds, due to the employment of reduced pictures.

As the quality of the foreground objects is quite reasonable, additional subjective tests were carried out on the "Snow" sequence. Three different decoded videos on the "Snow" are prepared; one is the result of the normal background separation coding, the other two are picture enhanced videos [4], where the enhancement is carried out for both foreground and background videos, and for only a foreground video, respectively. Again, the 7 Mbps H.264 is used as a reference. The question of "Which do you feel better for a surveillance purpose?" is given to every observer before the evaluation tests. This time, background separation approach always gets higher scores than the H.264 reference as shown in Fig. 5(b). However, the enhancement of both foreground and background videos does not get higher scores. The enhancement of only the foreground video results in the highest score. This fact can be considered that people like to see a normal scenery, but once the surveillance task is assigned to the observers, they like to see the enhanced objects for easy observation.

Let us compare the proposed approach with that, using GMM. The video of "Snow" in Fig. 1(c) shows the effectiveness of the STGMM employment. As only a car is running in the heavy snow in STGMM, the foreground bit-rate of 2.0 Mbps is enough for the proposed approach. In contrast, the GMM approach detects many snow falls as foreground objects which can be seen in Fig. 1(b). Therefore, the GMM approach makes busy foreground frames. The difference of bit-rates among the standard H.264, the proposed approaches using GMM and STGMM on the "SNOW" video are summarized in Fig. 6(a). The bit-rate of the GMM approach shows always the total maximum bit-rate



(a) Bit Rate Comparison (b) PSNR Comparison

Fig.6. PSNR and Bit Rate comparison: H.264 encoder and background separation coding approaches by using GMM and STGMM are compared.

of 3.5 Mbps. In addition, the PSNR of the running car is measured among these encoders in Fig. 6(b). The proposed approach shows about 10 dB higher PSNR than the H.264, although the bit-rate is less than a half. However, the GMM approach shows less than 15 dB performance than the H.264.

For further bit rate reduction, the chroma-key approach can be replaced by other efficient coding approaches. For example, when foreground information on every 16x16 blocks is encoded by the Modified Huffman coding, it results in 70 kbps in the Cross video which is around 1/3 smaller than the 0.2 Mbps in the blue only TS stream. However, it requires some special transmission format in TS stream. Therefore, the chroma-key approach has been employed for practical use. Although blue dots appear in some foreground objects by chroma-key approach, these dots can be easily removed by the 16x16 block based chroma-key control.

V. Conclusion

A background separation encoding approach has been proposed, by using the STGMM segmentation. For the surveillance purpose, around half a bit-rate employment is shown to perform better object PSNR. In addition, the picture enhancement on only the foreground areas is shown to be better than the 7 Mbps H.264 in subjective tests

REFERENCES

[1] "Special Issue on MPEG-4", IEEE Trans. on CAS for VT, Vol. 7 No.1 1997

[2] T. Nishi, et al., "Object-based Video Coding using Pixel State Analysis", IEEJ Trans. EIS, Vol.124, No.12, 2004.

[3] R. Ding et al., "Background-frame Based Motion Compensation for Video Compression" Proc. of ICME, 2004.

[4] Hiroaki Tezuka and Takao Nishitani, "Multiresolutional Gaussian Mixture Model for Precise and Stable Foreground Segmentation in Transform Domain", IEICE Trans. Fundamentals, Vol. E92-A, No.3,pp772-778, March 2009.

[5] C.Stauffer and W.E.L.Grimson, "Adaptive background mixture models for real-time tracking," Proc. CVPR'99, pp.246–252, 1999.

[6] Takeshi Okuno and Takao Nishitani, "Efficient Multi-Scale Retinex Algorithm Using Multi-Rate Image" to appear in the proceeding of ICIP 2009.