

Improving the Coding of Regions of Interest

Yi-Lun Lin, Shu-Fa Lin, and Homer H. Chen
Graduate Institute of Communications Engineering,
National Taiwan University
Taipei, Taiwan
Email: {b90901018, r92942088}@ntu.edu.tw,
homer@cc.ee.ntu.edu.tw

Yuh-Feng Hsu
CCRL, Industrial Technology Research Institute
Chutung, HsinChu, Taiwan
Email: yfhsu@itri.org.tw

Abstract—This paper considers a video coding system for surveillance applications. It consists of one "base encoder" that encodes a down-sampled, full-view version of the input video sequence and one "region of interest" (ROI) encoder that encodes an ROI of the video sequence at the original image resolution. An important requirement of the video coding system is that the ROI bitstream and the base bitstream should be independently decodable. We explore the inter-relationship between the full-view video sequence and the ROI video sequence and apply it to improve the computational efficiency of the ROI encoder. The proposed algorithm achieves an average of 170% speedup.

I. INTRODUCTION

In the last decade, many useful video coding algorithms have been developed and adopted in various video coding standards such as H.263 and MPEG-4. Video signals are compressed before storage or transmission over networks by exploiting spatial and temporal redundancies. High data compression efficiency brought by these digital video coding standards makes possible the wide deployment of video surveillance systems.

Most video surveillance systems today provide full-view videos. However, for security reasons, it is often required to show some specific regions of a video in more details than the other parts of the video. Such regions are often called regions of interest (ROI). In a conventional surveillance system, one single (fixed or pan-tilt-zoom) camera cannot provide both the full-view video and the ROI video at the same time because they have different resolution requirements. High resolution is preferable for the ROI while low resolution is sufficient for the full-view video. Although using two cameras simultaneously may work, the cost is doubled. Besides, the synchronization between cameras is not a trivial issue.

In the system described in this paper, we use one high-definition camera to capture the video data and produce

(through down-sampling) a full-view video at low resolution and an ROI video at the original resolution. Another requirement of the system is that, unlike the conventional scalable video coding schemes, each of the compressed video bitstreams should be independently decodable. For this reason, each of the full view video and the ROI videos needs one encoder, as shown in Figure 1. The architecture of the surveillance system under consideration is shown in Figure 2. The system consists of a base encoder to encode the down-sampled, full-view video and an ROI encoder to encode ROI sequence at the original camera resolution. Note that there may be more than one ROI in a video sequence. For simplicity, only one ROI is considered in this paper. However, the algorithm can be applied to dealing with multiple ROIs.

Since the ROI sequence and its corresponding region in the base sequence have the same content but at different resolutions, it is expected that certain spatial correlations exist between these two sequences. Therefore, we can exploit such correlations in the encoding process to save the computational cost. Our goal is to reduce the computational complexity of the ROI encoder while maintaining its output video quality.

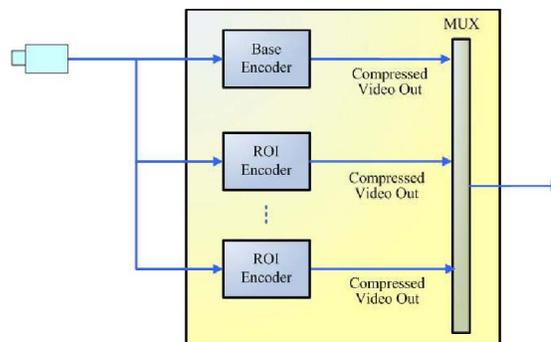


Figure 1. The base and ROI encoders and a high-resolution camera.

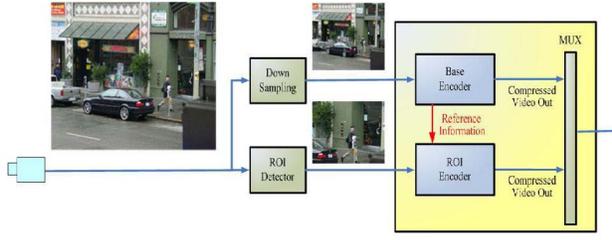


Figure 2. System architecture. We consider only one ROI encoder in this paper.

This paper is organized as follows. The analysis of the mode correlation between the base and the ROI sequences is given in Section II. Section III describes the proposed method for fast mode decision in detail. Section IV shows the simulation results, followed by a conclusion in Section V.

II. MODE ANALYSIS

A. Inter mode analysis

Intuitively, it should be highly likely that the best coding modes chosen by the ROI encoder and by the base encoder for corresponding blocks are the same. That is, when the base encoder encodes a macroblock with, say, the SKIP mode, the probability that the ROI encoder also encodes the corresponding macroblocks with the SKIP mode should be high.

TABLE I. MODE DISTRIBUTION (%)

ROI \ Base	SKIP	16x16	16x8	8x16	P8x8	I4MB	I16MB
SKIP	76.4	13.1	3.6	3.5	2.7	0.1	0.7
16x16	24.2	34.8	9.7	9.1	20.7	0.8	0.8
16x8	21.9	32.2	13.0	10.3	19.9	1.6	1.1
8x16	22.6	31.6	10.6	12.3	19.8	1.8	1.2
P8x8	7.2	27.2	10.9	10.5	41.4	2.2	0.3
I4MB	6.0	11.1	6.6	6.6	10.0	51.2	8.5
I16MB	48.1	11.6	3.3	3.2	1.4	6.6	25.8

To verify our conjecture, we tests 8 sequences (*Silent*, *Tempete*, *Flower*, *Container*, *Foreman*, *Mobile*, *Stefan*, and *Mother & Daughter*) with H.264 reference software JM 8.5 [1]. The ROI sequences are in CIF resolution, while the base sequences are in QCIF resolution.

The spatial relation between the base sequence and the ROI sequence is shown in Figure 3. The resulting coding modes determined by the JM 8.5 reference software is shown in Table I, where P8x8 denotes the set of modes {8x8, 8x4, 4x8, and 4x4}, I4MB denotes the Intra 4x4 mode, and I16MB denotes the Intra 16x16 mode. Here, for each mode chosen by the base encoder, we show the distribution of the best modes determined by the ROI encoder. As an example, the 2nd row indicates the mode distribution for ROI blocks that correspond to the base blocks of the 16x16 mode. The

first entry in this row indicates that 24.2% of such ROI blocks are coded with the SKIP mode by the ROI encoder. Likewise, when the base encoder chooses the SKIP mode for a block, the probability that the ROI encoder also chooses the SKIP mode for the corresponding ROI block is 76.4%. Note that the probability that both encoders choose the Intra 4x4 mode is 51.2% and the P8x8 mode 41.4%. These results support our conjecture and suggest that the coding modes for ROI blocks can be predicted from their corresponding blocks in the base sequence.

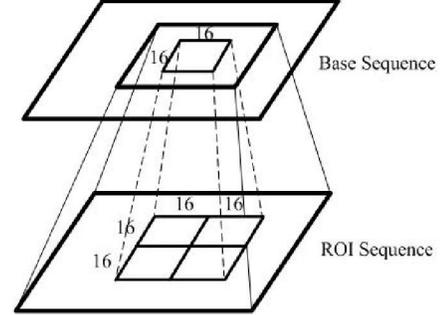


Figure 3. Corresponding region

Moreover, we collect the mode information from all the base sequences and find that 40% of the macroblocks in the base sequences are coded with the SKIP mode, 25% of them are coded with the 16x16 mode, and 20% are coded with the P8x8 mode. As mentioned before, the probability that both the base encoder and the ROI encoder choose from any one of these three modes is high. Therefore, for more than half of the macroblocks in the ROI sequence, the best mode is likely to be predicted correctly.

By predicting the best mode, we can greatly reduce the computations in the ROI encoder and, if the best mode can be predicted correctly, we can also maintain the video quality. The proposed algorithm uses multiple modes as the prediction candidates to improve the prediction accuracy to achieve this goal. The detail is described in Section III.

B. Intra-prediction mode analysis

Since high correlation exists between the best inter mode of the base encoder and of the ROI encoder, we expect certain correlation also exists between the best intra- modes of the two encoders.

The Intra4x4 mode distribution is shown in Table II. As expected, it is highly likely that both the two encoders choose the same intra mode. Besides, we also note that whether the best intra mode of the base encoder is, the probability that mode 2 (DC mode) to be chosen as the best prediction mode is always high in the ROI encoder. Therefore, not only the best intra mode in the base sequence for the corresponding blocks would be included as a prediction candidate, but also the DC mode.

Table III shows the Intra16x16 mode distribution, which is similar to Table II. We can also observe that the mode 2 (DC mode) always has good chance to be the best one.

C. Chroma-prediction mode analysis

When a macroblock is going to be encoded as an intra block, it has good chance that the macroblock chooses DC mode as the best chroma mode. As shown in the Table IV, whether the best chroma mode of the base encoder is, the probability that DC mode (mode 0 in the chroma modes) is chosen as the best mode is always high. In addition, the probability that the two encoders choose the same chroma mode is also higher than the other ones except the DC mode.

TABLE II. INTRA4x4 PREDICTION MODE DISTRIBUTION (%)

ROI \ Base	0	1	2	3	4	5	6	7	8
0	38.0	4.1	5.3	10.3	8.3	16.7	6.5	17.3	6.6
1	6.2	46.6	7.4	10.1	10.8	8.5	20.8	8.6	23.2
2	35.1	30.2	73.5	30.1	25.6	26.0	24.0	25.1	24.3
3	3.7	2.3	2.6	18.6	4.5	5.0	4.1	12.2	7.8
4	3.0	2.7	2.2	4.9	21.1	10.8	10.4	5.1	5.3
5	4.33	1.9	1.9	4.5	9.3	17.2	5.0	5.2	4.0
6	2.47	5.0	2.5	4.9	10.8	5.9	19.1	4.6	6.4
7	4.8	1.7	2.0	8.9	4.2	5.2	3.5	15.4	5.1
8	2.5	5.6	2.6	7.8	5.5	4.8	6.5	6.3	17.3

TABLE III. INTRA16x16 PREDICTION MODE DISTRIBUTION (%)

ROI \ Base	0	1	2	3
0	22.2	17.5	16.9	21.2
1	23.2	28.7	22.7	23.5
2	36.2	35.7	42.4	35.3
3	18.4	18.1	18.0	20.0

TABLE IV. CHROMA PREDICTION MODE DISTRIBUTION (%)

ROI \ Base	0	1	2	3
0	99.3	70.4	71.8	61.7
1	0.3	21.5	4.9	15.3
2	0.3	3.9	19.0	13.1
3	0.1	4.3	4.2	9.9

III. ALGORITHM DESCRIPTION

For each block type, more than one mode would be used as prediction to improve the prediction accuracy, as shown in Figure 4. The most possible mode is set as one of the prediction candidates for each mode. Usually, the most possible mode is the same as the best mode chosen by the base encoder. For example, when the corresponding region in the base sequence is encoded with the P8x8 mode, the

most possible mode is the P8x8 mode with 41.4% probability; when the corresponding region in the base sequence is encoded with the Intra4x4 mode, the most possible mode is the Intra4x4 mode with 51.2% probability, as shown in Table I.

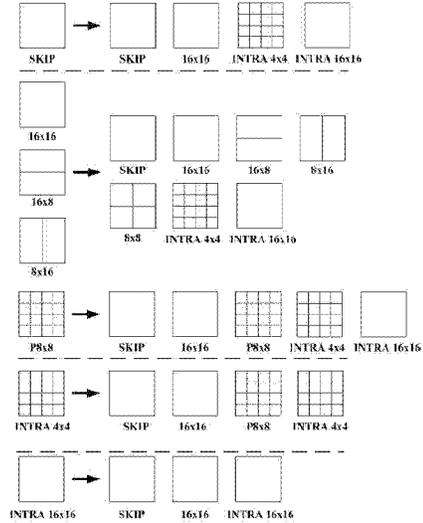


Figure 4. Mode prediction in the ROI encoder

Since the mode set {16x16, 16x8, and 8x16} do not have obvious tendency, the proposed algorithm check all the modes as usual. Besides, as shown in Table V, when the base encoder chooses the mode set {16x16, 16x8, and 8x16} and the ROI encoder chooses the P8x8 mode, more than 50% of the macroblocks in the ROI sequences would be encoded with the 8x8 mode. Therefore, instead of the mode set {8x8, 8x4, 4x8, 4x4}, only the 8x8 mode is chosen as a prediction candidate in this case. Note that the SKIP mode and the 16x16 mode are frequently chosen as the best mode. Therefore, these two modes are always included as the prediction candidates. Also note that the P8x8 mode is one of the prediction candidates for the Intra 4x4 mode, because both modes are often used to encode a complex region.

TABLE V. MODE DISTRIBUTION (%) WHEN THE P8x8 MODE IS THE BEST MODE IN THE ROI ENCODER

ROI \ Base	8x8	8x4	4x8	4x4
16x16	52.2	22	18.7	7.1
16x8	52.3	19.5	21.1	7.1
8x16	50.6	20.8	20.3	8.3

The Intra16x16 mode and the Intra4x4 mode are used to encode the regions containing new objects and have great influence on video quality; therefore, the two intra modes are always included as the prediction candidates in the ROI encoder. According to the analysis in the previous section, we propose to turn on two Intra-prediction modes only, which are the DC mode and the prediction mode chosen by the base encoder in the corresponding region, instead of

turning on all the Intra-prediction modes (9 modes in Intra4x4 and 4 modes in Intra16x16). We also apply the same rule to Chroma prediction. By reducing the prediction modes, the proposed algorithm reduces the RDO calculation from 592 to only 68 and saves a lot of encoding time.

In the proposed algorithm, we predict the most possible modes in the ROI based on the mode information of the corresponding macroblocks encoded by the base encoder. Each blocktype has more than 80% accuracy, and thus we can improve the computational efficiency of the ROI encoder while maintaining the output video quality. The proposed mode prediction requires only little memory overhead for saving extra mode information of the corresponding regions in the base sequence.

IV. SIMULATION RESULTS

The proposed algorithm is implemented in the H.264 reference software JM 8.5. In our simulation, the search range of motion estimation is set to ± 16 pels for QCIF resolution and ± 32 pels for CIF resolution. Apart from the first frame, all the other frames are encoded as P frames. Furthermore, R-D optimization, Hadamard transform, and CAVLC entropy coding are turned on for all the tests. All block modes and one reference frame are used for both encoders. The detailed simulation conditions used in H.264 reference software are shown in Table VI. The simulations are performed on a PC with a Pentium IV, 2.4 GHz CPU and a 768 MB RAM.

TABLE VI. SIMULATION CONDITIONS

Frame Rate	30 fps
Hadamard Transform	Used
Search Range	16 for QCIF/ 32 for CIF
Sequence Type	IPPP
Entropy Coding Method	CAVLC
RD-Optimized Mode Decision	Used
QP	28, 32, 36, 40

We test various sequences to evaluate the performance of the proposed algorithms. To present the test result in a more compact form, we use Bjontegaard delta PSNR (BDPSNR) and Bjontegaard delta bit rate (BDBR) [9] instead of the RD-plots. As shown in Table VII, the proposed algorithm achieves 1.69 times speedup on average. By the proposed mode prediction, the encoding time is greatly saved. However, the computational complexity of the motion search for the predicted modes is still high. To further speed up the encoding of the ROI, some fast motion estimation techniques such as those described in [5]-[8] and the early termination criteria in [3]-[5] can also be applied.

V. CONCLUSION

In this paper, we have described a multiple-encoder video coding system for surveillance applications. The system captures video data using a high-definition video camera and

decomposes the video data into one full-view but low resolution sequence and one full-resolution ROI sequence. The two video sequences are encoded separately to allow independent decoding so that any data loss of the base sequence during transmission will not affect the video quality of the ROI sequence. The ROI encoder we have developed is able to reduce the encoding time to half. The computational efficiency of the ROI encoder is accomplished by exploiting the inter-relationship between the ROI sequence and the full-view sequence. Although we only consider one ROI encoder in this paper, the method is readily applicable to cases where more than one ROI encoder are employed.

TABLE VII. PSNR COMPARISON

Sequence (CIF)	BDPSNR (dB)	BDBR (%)	Encoding Speedup
Mobile	-0.47	4.04%	1.66
Tempete	-0.15	1.73%	1.74
Flower	-0.21	2.73%	1.68
M & D	-0.15	1.83%	1.66
Container	-0.15	1.87%	1.67
Silent	-0.15	1.68%	1.77
Foreman	-0.17	2.48%	1.74
Stefan	-0.41	4.26%	1.67
Average	-0.23	2.58%	1.70

REFERENCES

- [1] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification* (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), 2003.
- [2] D. Wu, F. Pan, K. P. Lim, S. Wu, Z. G. Li, X. Lin, S. Rahardja, and C. C. Ko, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 953-958, Jul. 2005.
- [3] C.-Y. Chang, C.-H. Pan, and H. H. Chen, "Fast mode decision for P-frames in H.264," *Picture Coding Symposium*, Dec. 2004.
- [4] S.-F. Lin, C.-Y. Chang, C.-C. Su, Y.-L. Lin, C.-H. Pan, and H. H. Chen, "Fast multi-frame motion estimation and mode decision for H.264 encoders," *International Conf. on Wireless Networks, Communications, and Mobile Computing*, Maui, Hawaii, Jun. 2005.
- [5] P. Yin, H.-Y. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for H.264," in *Proc. IEEE Int. Conf. on Image Processing*, vol. III, pp. 853-856, Sept. 2003.
- [6] Z. Chen, P. Zhou, and Y. He, "Fast motion estimation for JVT," JVT-G016.doc, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, 7th Meeting, Mar. 2003.
- [7] K. K. Ma and G. Qiu, "Unequal-arm adaptive rood pattern search for fast block-matching motion estimation in the JVT/H.26L," in *Proc. IEEE Int. Conf. on Image Processing*, Sept. 2003.
- [8] S.-F. Lin, M.-T. Lu, C.-H. Pan, and H. H. Chen, "Fast multi-frame motion estimation for H.264 and its applications to complexity-aware streaming," in *Proc. IEEE Int. Symposium on Circuits and System*, pp. 1505-1508, May 2005.
- [9] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Doc. VCEG-M33, Apr. 2001.
- [10] Feng Pan, Xiao Lin, Susanto Rahardja, Keng Pang Lim, Z. G. Li, Dajun Wu, and Si Wu, "Fast Mode Decision Algorithm for Intraprediction in H.264/AVC Video Coding," *IEEE Transactions on Circuits and System*, vol. 15, July 2005