Repeat-Frame Selection Algorithm for Frame Rate Video Transcoding

Yi-Wei Lin m9723030@ems.ndhu.edu.tw
Gwo-Long Li m9323004@ems.ndhu.edu.tw
Mei-Juan Chen cmj@mail.ndhu.edu.tw
Department of Electrical Engineering
National Dong-Hwa University
Hualien, 97401 Taiwan, R.O.C.

Chia-Hung Yeh*(Corresponding author) yeh@mail.ee.nsysu.edu.tw
Department of Electrical Engineering
National Sun Yat-Sen University
Kaohsiung, 80424 Taiwan, R.O.C.

Shu-Fen Huang m9823002@ems.ndhu.edu.tw
Department of Electrical Engineering
National Dong-Hwa University
Hualien, 97401 Taiwan, R.O.C.

Abstract

To realize frame rate transcoding, the forward frame repeat mechanism is usually adopted to compensate the skipped frames in a video decoder for end-device. However, based on our observation, it is unsuitable for repeating all skipped frames only in the forward direction and sometimes the backward repeat may provide better results. To deal with this issue, we propose a new reference frame selection method to determine the direction of repeat-frame for skipped Predictive (P) and Bidirectional (B) frames. For P-frame, the non-zero transformed coefficients and the magnitude of motion vectors are taken into consideration to determine the use of forward or backward repeat. For B-frame, the magnitude of motion vector and its corresponding reference directions of the blocks in B-frame are selected as the decision criteria. Experimental results show that the proposed method provides 1.34 dB and 1.31 dB PSNR improvements in average for P and B frames, respectively, compared with forward frame repeat.

Keywords: Transcoding, Temporal transcoding, Frame-rate transcoding, Frame skipping, Forward/Backward repeat.

1. INTRODUCTION

In recent years, the applications of multimedia [1]-[4] are rising and popular. One of applications, video transcoding becomes an important issue in video communication with the development of the network transmission. Video transcoder converts videos into different qualities, frame rates, resolutions, even the coding standards [5]-[8] to fit the network variation. The concept of video transcoding [9] is shown in FIGURE 1. When network bandwidth is insufficient, three kinds of methods can be used to convert a bitstream into different bitrates and they are quality transcoding, spatial transcoding and temporal transcoding that is also named as frame rate transcoding. For quality transcoding [10]-[12], the quantization parameter (QP) is adjusted in encoder to fit target bitrate under the bitrate constraint. In addition, another way to achieve video transcoding is to adjust the spatial resolution of a sequence for transmission purpose [13]-[19]. In spatial transcoding, the shrunk sequence saves the bitrate and the decoder recovers the sequence to the original size when receiving data. However, the way to choose the down-scaling and up-scaling method is still a great challenging issue. Sometimes, we adjust the frame rate of a sequence to fit the target bitrate.

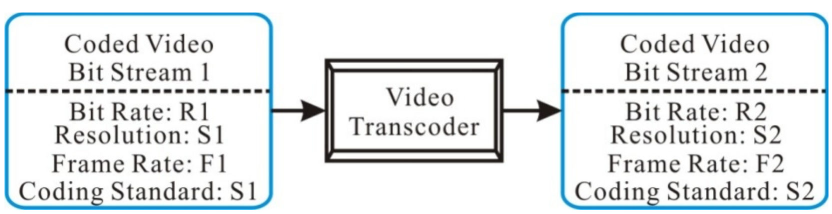
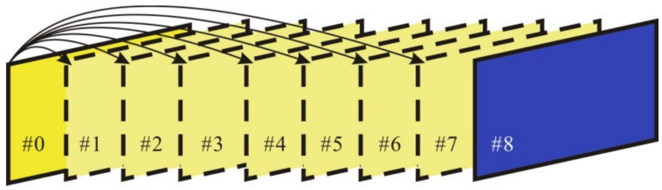
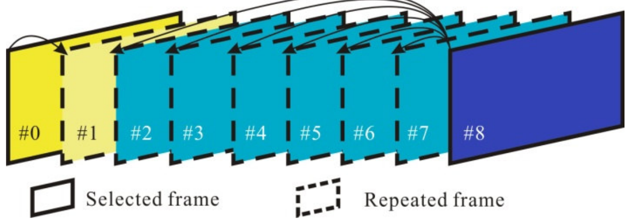


FIGURE 1: Concept of Video Transcoding

Many methods have been proposed for temporal transcoding and most of them focus on two directions, frame rate decision and key frame selection [20]-[28]. That is, we should decide the acceptable frame rate according to the current bandwidth and select the most significant frames in a group of picture (GOP). After that, a traditional video decoder repeats the previous frame to compensate skipped frames shown in FIGURE 2 (a) for GOP=8 and we call it as "Regular Forward Repeat Method" (RFRM). In lots of experiments, we observe that repeating all skipped frames in the forward direction seems not appropriate and backward repeat as shown in FIGURE 2 (b) may have better results.



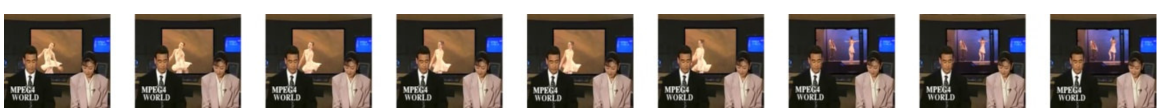
(a) Regular Frame Repeat Method



(b) Repeat-Frame Selection Methods

FIGURE 2: An Example for Sequence in GOP=8

FIGURE 3 shows an example of the benefit of backward frame repeat. FIGURE 3 (a) shows the decoded sequence without dropping frames and FIGURE 3 (b) is transcoded by frame rate descending with forward repeat, and FIGURE 3 (c) shows dynamic forward or backward frame repeat. From this example, it can be seen that the results after temporal transcoding can be improved significantly by considering both forward and backward repeats. Based on the observation mentioned above, this paper proposes a reference frame repeat method to determine the direction of repeat-frame for skipped frames during transcoding process. For P-frame, the number of nonzero transformed coefficients and the magnitude of motion vectors are jointly considered to determine the repeat direction. For B-frame, the prediction directions and the magnitude of motion vectors are combined to obtain the criteria for the repeat direction determination. This paper is organized as follows. In Section 2, the proposed "Repeat-Frame Selection Methods" (RFSM) is explicated. Section 3 presents the extensive experimental results to verify the efficiency of our methods. Finally, concluding remarks are given in Section 4.



(a) Decoded Sequence without Dropping Frames

(b) Repeat Forward Frame Regularly by RFRM

(c) Dynamic Frame Repeat

#144 #145 #146 #147 #148 #149 #150 #151 #152

FIGURE 3: An Example of Different Methods in *News* Sequence for GOP=8

2. PROPOSED METHODS

FIGURE 4 shows the system diagram for the proposed system. We propose the RFSM to determine that the skipped frames should be repeated by either forward or backward direction. Instead of fully reconstructing pixel data, the proposed algorithm just employs the motion vector information, non-zero transformed coefficients and the prediction directions of B-frame partially decoded from the bitstream1 to determine the repeat direction. The Encoder2 in temporal transcoder will embed the results of forward/backward decision into Bitstream2 by inserting repeat direction into the headers. After receiving the Bitstream2, the Decoder2 can decompress the video bitstream and dynamic forward/backward frame repeat can be executed. The proposed methods for P-frame and B-frame cases are discussed in the following two subsections separately.

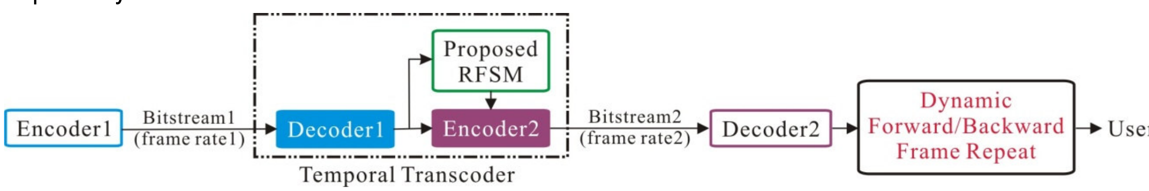


FIGURE 4: Diagram of Proposed Repeat-Frame Selection Method

2.1 P frames

For P-frame, since only the information of forward direction is available, the factors we consider are the magnitude of motion vectors and the number of non-zero transformed coefficients. Normally, the motion activity and the number of non-zero transformed coefficients indicate the property of the sequence and the complexity of the frame, respectively. From our observation, the high-motion sequence results in obvious amount of non-zero transformed coefficients. In our proposal, we define a selective factor (SF) as follows.

$$SF_P = \sum_{i=1}^{N} (NZcoeff_i \times MV_i) , \tag{1}$$

$$MV_i = \sum_{k=1}^{M_i} (|MVX_k| + |MVY_k|) , \tag{2}$$

where N refers to the number of the macroblock in one frame, M_i refers to the number of blocks in the i-th macroblock, MV_i refers to the sum of the motion vector magnitude in X and Y directions and $NZcoeff_i$ refers to the number of non-zero transformed coefficients in the i-th macroblock. After the SF_P of each skipped frame is calculated, we select the frame with the maximum SF_P as the separated frame which means two consecutive frames have higher motion activity variation shown in FIGURE 5. Finally, the frames after the separated frame (including the separated frame) in a GOP are assigned as backward repeat. Take FIGURE 2(b) for example, the frame #2 is the separated frame in a GOP. In FIGURE 5, the frame #258 is the separated frame.

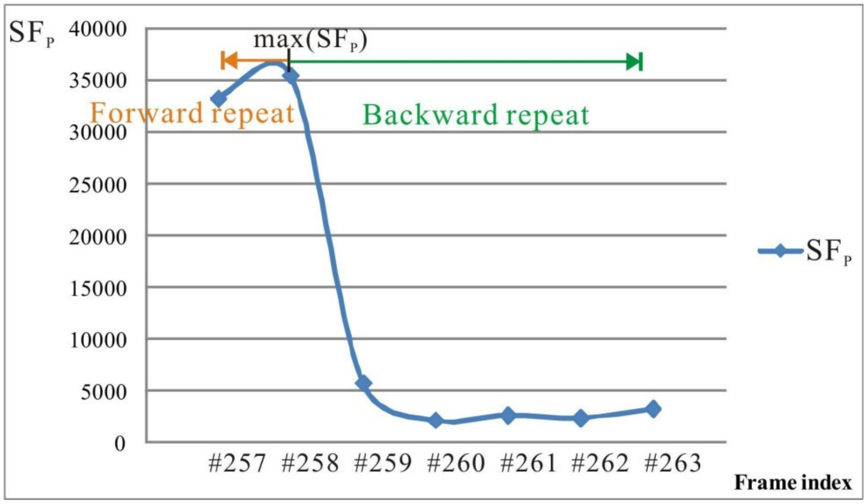


FIGURE 5: Separated Frame Determination of *Foreman* Sequence for P-Frame in GOP=8 case

2.2 B frames

For B-frame, we first decode the bitstream to obtain the magnitude of motion vectors and the prediction directions of each block. We avoid reconstructing pixel values in order to reduce computational complexity. It is well-accepted that the larger magnitude of motion vector implies the higher motion activity or scene change that the frame may contain. Therefore, the magnitude of motion vectors is selected as a factor and its prediction direction of decoded blocks is also included in the proposed method. If a block is encoded as forward prediction mode, it means that the most similar block is forward prediction rather than backward. Oppositely, backward prediction mode implies that the best match block can be found from the following frames. As a result, we take the magnitude of the motion vector in Forward or Backward prediction to be the factor of separated frame determination. SF_B is defined as the difference of two factors, $MV_{Forward}$ and $MV_{Backward}$, to stand for the motion tendency of forward and backward prediction directions in a frame.

$$SF_B = MV_{Backward} - MV_{Forward}\ , \tag{3}$$

$$MV_{Forward} = \sum_{u=1}^{U} (|MVX_{Forward\_u}| + |MVY_{Forward\_u}|)\ , \tag{4}$$

$$MV_{Backward} = \sum_{v=1}^{V} (|MVX_{Backward\_v}| + |MVY_{Backward\_v}|)\ , \tag{5}$$

where *MVX* and *MVY* represent the magnitude of motion vectors in *X* and *Y* directions for each block, respectively, and *U and V* are the numbers of blocks in forward and backward directions, respectively, of one frame. Finally, we select the frame with the maximum SF_B to be the separated frame in a GOP shown in FIGURE 6. Once the separated frame is determined, the frames after the separated frame (including separated frame) in a GOP are assigned as backward repeat.

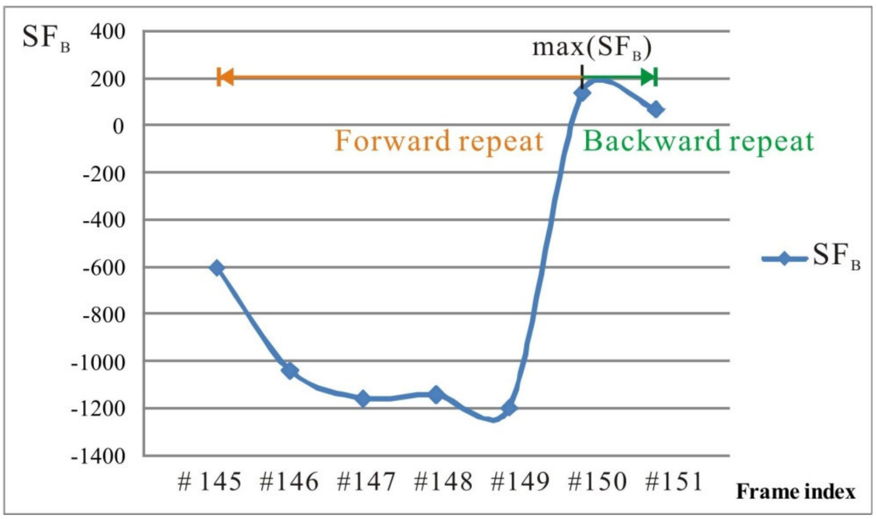


FIGURE 6: Separated Frame Determination of *News* Sequence for B-Frame in GOP=8 case

3. EXPERIMENTAL RESULTS

In this section, we compare the proposed method with *SAD-based frame repeat* method and RFRM in terms of subjective and objective qualities to demonstrate the efficiency of our method. *SAD-based* method decodes all frames to the pixel-domain and calculates the Sum of the Absolute Difference (SAD) between the forward and the backward reference frames. Afterwards, SAD is used to determine the forward and backward repeat. In *SAD-based* method, the skipped frame which first satisfies the condition of the backward SAD less than forward SAD will be selected as the separated frame. All methods are implemented on the H.264 JM 15.1 [29] reference software. The simulation setting is in the following. The test benchmark sequences include *CarPhone*, *Foreman*, *Mobile*, *News*, *Salesman* and *Silent* in QCIF resolution with 289 frames and the search range is 16. The coding structures are IPPP in P-frame and IBBBP in B-frame for GOP=4 and so on.

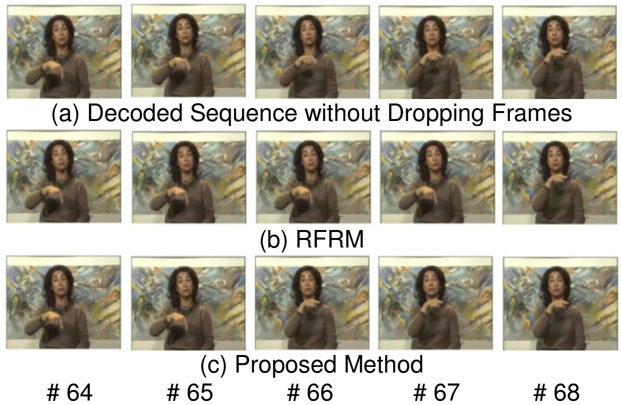


FIGURE 7: Subjective Comparison of *Silent* Sequence for P-Frame in GOP=4

FIGURE 7 to FIGURE 11 show the subjective quality comparisons for P-frame and B-frame of RFRM and our method. Take FIGURE 8 as example, the worker's hand waved out the scene in #258. Our proposed method selects the proper frame to repeat. It is evident that our proposed algorithm is very similar to the original sequence.

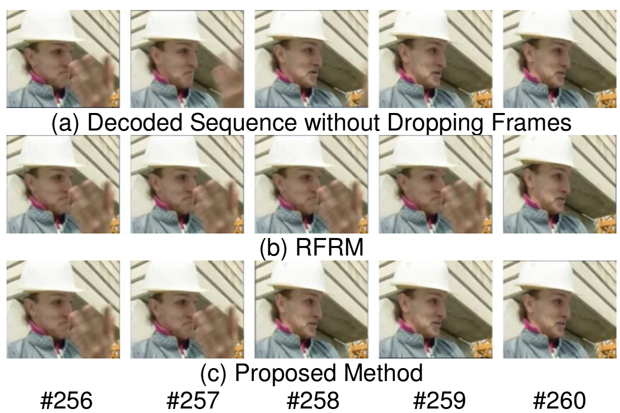


FIGURE 8: Subjective Comparison of *Foreman* Sequence for B-Frame in GOP=4

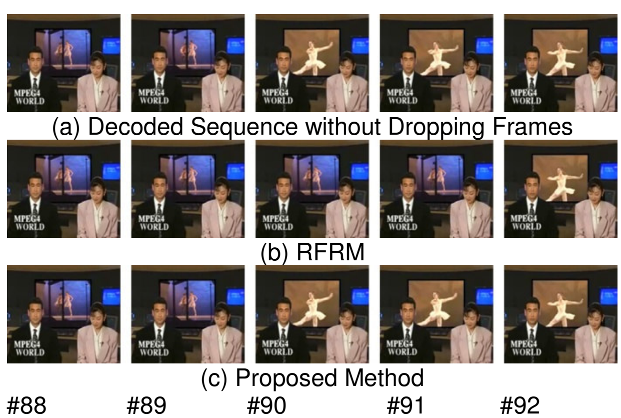


FIGURE 9: Subjective Comparison of *News* Sequence for B-Frame in GOP=4

FIGURE 10: Subjective Comparison of *CarPhone* Sequence for P-Frame in GOP=6

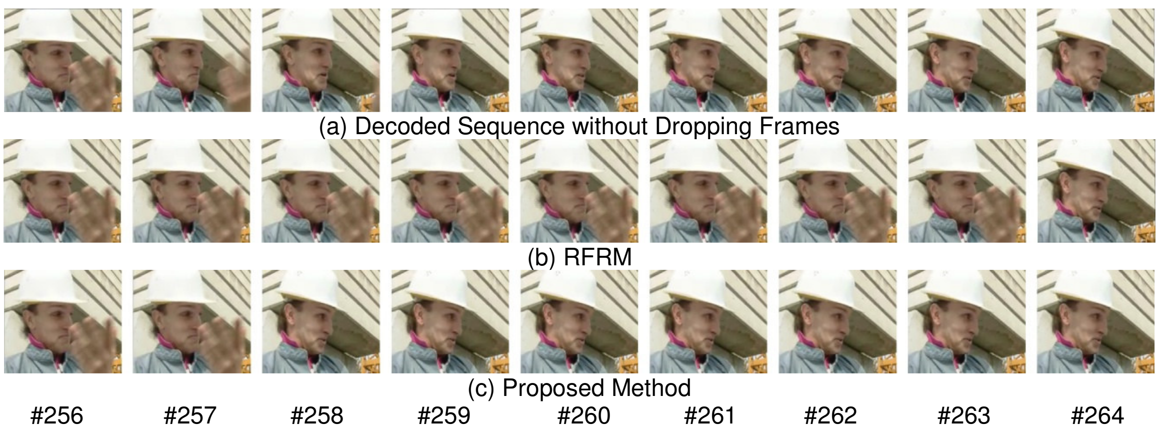


FIGURE 11: Subjective Comparison of *Foreman* Sequence for P-Frame in GOP=8

	GOP 4				**GOP 6**				**GOP 8**			
	SAD-based method	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based method*	*RFRM*	*Proposed*	*Proposed-RFRM*
CarPhone	24.04	22.70	23.63	0.93	25.72	24.15	25.27	1.12	26.54	24.45	25.86	1.41
Foreman	20.93	19.03	20.34	1.31	21.92	19.63	21.24	1.61	22.21	19.54	21.76	2.22
Mobile	17.87	16.04	17.22	1.18	18.43	16.18	17.69	1.51	18.11	15.95	17.11	1.16
News	23.16	21.85	22.66	0.81	25.18	23.61	24.30	0.69	24.87	23.21	24.22	1.01
Salesman	28.25	26.55	27.88	1.33	30.35	28.41	29.84	1.43	30.95	28.83	30.70	1.87
Silent	24.27	22.84	24.09	1.25	26.19	24.40	25.87	1.47	26.72	24.69	26.46	1.77
Average	**23.09**	**21.50**	**22.64**	**1.14**	**24.63**	**22.73**	**24.04**	**1.31**	**24.90**	**22.78**	**24.35**	**1.57**

TABLE 1: PSNR Comparison for P-Frame (dB)

	GOP 4				GOP 6				GOP 8			
	SAD-based method	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based* method	*RFRM*	*Proposed*	*Proposed-RFRM*	*SAD-based* method	*RFRM*	*Proposed*	*Proposed-RFRM*
CarPhone	24.25	22.79	23.95	1.16	25.95	24.2	25.60	1.40	26.68	24.51	26.34	1.83
Foreman	21.04	19.08	20.77	1.69	22.00	19.67	21.58	1.91	22.33	19.60	21.98	2.38
Mobile	18.10	16.21	16.87	0.66	18.63	16.32	16.94	0.62	18.30	16.08	16.47	0.39
News	22.66	21.91	22.42	0.51	25.01	23.68	24.04	0.36	25.04	23.31	24.63	1.32
Salesman	28.09	26.95	28.15	1.20	30.47	28.53	30.28	1.75	30.82	28.70	30.59	1.89
Silent	24.77	23.19	24.46	1.27	26.73	24.78	26.21	1.43	26.85	24.69	26.42	1.73
Average	**23.15**	**21.69**	**22.77**	**1.08**	**24.80**	**22.86**	**24.11**	**1.25**	**25.00**	**22.82**	**24.41**	**1.59**

TABLE 2: PSNR Comparison for B-Frame (dB)

	P Frame		B Frame	
	RFRM	***Proposed***	***RFRM***	***Proposed***
GOP 4	75.1%	74.5%	72.6%	72.4%
GOP 6	82.9%	82.3%	81.1%	80.7%
GOP 8	87.1%	86.4%	85.2%	84.9%

TABLE 3: Decreased Computational Complexity compared with *SAD-based* Method

TABLE 1 and TABLE 2 show the PSNR comparisons for P-frame and B-frame. In those tables, we found that the PSNR improvements of our method is 1.14 dB, 1.31 dB and 1.57 dB for GOP=4, GOP=6 and GOP=8, respectively, when compared with RFRM method for P-frame case, and 1.08 dB, 1.25 dB and 1.59 dB for GOP=4, GOP=6 and GOP=8, respectively, when compared with *RFRM* method for B-frame case. Furthermore, when compared with *RFRM* method, the average PSNR improvements of all GOP sizes are 1.34 dB and 1.31 dB for P-frame and B-frame cases, respectively. TABLE 3 shows the decreased computational complexity, which is measured by CPU time, compared with *SAD-based* method. *SAD-based* method needs to decode bitstream to reconstruct pixel values for SADs calculation while RFRM and our proposal need not, which save much encoding time. Our algorithms only utilize the magnitude of motion vectors, non-zero transformed coefficients and motion compensation directions to determine frame repeat direction. Although our proposed method increases the computational complexity slightly when compared with RFRM, the quality can be increased significantly.

4. CONCLUSIONS

In this paper, in place of the traditional regular forward repeat method for frame-rate transcoding, we propose efficient algorithms which can dynamically select suitable frame to repeat. Our experimental results show that the proposed method has significant PSNR improvements compared with traditional forward repeat. The proposed method can select the proper frame to repeat and achieve better subjective quality.

5. REFERENCES

1. N. A. Lili and K. Fatimah. "*Content modelling for human action detection via multidimensional approach*". International Journal of Image Processing, 3(1):17-30, 2009
2. F. Rohmad and Abdul Azim Abd Ghani. "*Empirical evaluation of decomposition strategy for wavelet video compression*". International Journal of Image Processing, 3(1):31-54, 2009
3. Y. Tabii and R.O.H. Thami. "*A framework for soccer video processing and analysis based on enhanced algorithm for dominant color extraction*". International Journal of Image Processing, 3(4):131-142, 2009
4. M.C. Chi, C.H. Yeh and M.J. Chen. "*Robust region-of-interest determination based on user attention model through visual rhythm analysis*". IEEE Transactions on Circuits and Systems for Video Technology, 19(7):1025-1038, 2009
5. H.F. Shen, X.Y. Sun and F. Wu. "*Fast H.264/MPEG-4 AVC transcoding using power-spectrum based rate-distortion optimization*". IEEE Transactions on Circuits and Systems for Video Technology, 18(6):746-755, 2008
6. A. Dziri, A. Diallo, M. Kieffer and P. Duhamel. "*P-picture based H.264 AVC to H.264 SVC temporal transcoding*". In Proceedings of IEEE International Wireless Communications and Mobile Computing Conference, 2008
7. X. Jing, W.C. Siu, L.P. Chau and A.G. Constantinides. "*Fast intra mode decision algorithm for H.263 to H.264/AVC transcoding*". In Proceedings of IEEE International Conference on Neural Networks and Signal Processing, 2008
8. M. Pantoja and N. Ling. "*Adaptive transform size and frame-field selection for efficient VC-1 to H.264 high profile transcoding*". In Proceedings of IEEE International Symposium on Circuits and Systems, 2009
9. J. Xin, C.W. Lin and M.T. Sun. "*Digital video transcoding*". Proceedings of the IEEE, 93(1):84-97, 2005
10. X. Xiu, L. Zhuo and L. Shen. "*A H.264 bit rate transcoding scheme based on PID controller*". In Proceedings of IEEE International Symposium on Communications and Information Technology, 2005
11. X.Y Wang, Y. Zhang, H.L. Li and W.L. Zhu. "*Adaptive rate control for dynamic bandwidth in video transcoding*". In Proceedings of IEEE International Conference on Communication Systems, 2008

12. Y.M. Zhou, Y. Sun, Z.D. Feng and S.X. Sun. "*New rate-complexity-quantization modeling and efficient rate control for H.264/AVC*". In Proceedings of 2008 IEEE International Conference on Multimedia and Expo, 2008
13. M.J. Chen, M.C. Chu and S.Y. Lo. "*Motion vector composition algorithm for spatial scalability in compressed video*". IEEE Transactions on Consumer Electronics, 47(3):319-325, 2001
14. P.Z., Y. Lu, Q. Huang and W. Gao. "*Mode mapping method for H.264/AVC spatial downscaling transcoding*". In Proceedings of IEEE International Conference on Image Processing, 2004
15. H. Sun and Y.P. Tan. "*Arbitrary downsizing video transcoding using H.26L standard*". In Proceedings of IEEE International Conference on Image Processing, 2003
16. H. Sun and Y.P. Tan. "*Fast motion re-estimation for arbitrary downsizing video transcoding using H.264/AVC standard*". IEEE Transactions on Consumer Electronics, 50(3):887-894, 2004
17. M. Pantoja and N. Ling. "*Transcoding with quality enhancement and irregular sampling*". In Proceedings of IEEE International Conference on Image Processing, 2008
18. H.Y. Shu and K. N. Ngan. "*Pre- and post-shift filtering for blocking removing in downsizing transcoding*". IEEE Transactions on Circuits and Systems for Video Technology, 19(6):882-886, 2009
19. X. Yu, E.H. Yang and H. Wand. "*Down-sampling design in DCT domain with arbitrary ratio for image/video transcoding*". IEEE Transactions on Image Processing, 18(1):75-89, 2009
20. M.A. Bonuccelli, F. Lonetti and F. Martelli, "*Temporal transcoding for mobile video communication*". In Proceedings of the 2nd Annual International Conference on Mobile and Ubiquitous System: Networking and Service, 2005
21. Y.H. Ho, W.R. Chen and C.W. Lin. "*A rate-constrained key-frame extraction scheme for channel-aware video streaming*". In Proceedings of IEEE International Conference on Image Processing, 2004
22. C.T. Hsu, C.H. Yeh, C.Y. Chen and M.J. Chen. "*Arbitrary frame rate transcoding through temporal and spatial complexity*". IEEE Transactions on Broadcasting, 55(4):767-775, 2009
23. J.N. Youn, M.T. Sun and C.W. Lin. "*Motion vector refinement for high-performance transcoding*". IEEE Transactions on Multimedia, 1(1):30-40, 1999
24. M.J. Chen, M.C. Chu and C.W. Pan. "*Efficient motion-estimation algorithm for reduced frame-rate video transcoder*". IEEE Transactions on Circuits and Systems for Video Technology, 12(4): 269-275, 2002

25. I.H. Shin, Y.L. Lee and H.W. Park. "*Motion estimation for frame-rate reduction in H.264 transcoding*". In Proceedings of 2nd IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, 2004
26. C.T. Hsu, C.H. Yeh and M.J. Chen. "*Effective frame rate decision by Lagrange optimization for frame skipping video transcoding*". In Proceedings of International Symposium on Visual Computing, 2008
27. K.C. Yang, G. Dane and K. El-Maleh. "*Temporal quality evaluation for enhancing compressed video*". In Proceedings of IEEE International Conference on Computer Communications and Networks, 2007
28. V. Chander, A. Reddy, S. Gaurav, N. Khanwalkar, M. Kakhani and S. Tapaswi. "*Fast and high quality temporal transcoding architecture in the DCT domain for adaptive video content delivery*". In Proceedings of IEEE International Conference Computer Engineering and Technology, 2009
29. Joint Video Team software JM15.1 http://iphome.hhi.de/suehring/tml/download/