

Correspondence

An Efficient Motion Vector Coding Scheme Based on Minimum Bitrate Prediction

Sung Deuk Kim and Jong Beom Ra

Abstract—Motion vector coding efficiency is becoming an important issue in low bitrate video coding because of its increasing relative bit portion. This work presents a new motion vector coding technique based on minimum bitrate prediction. In the proposed scheme, a predicted motion vector is chosen from the three causal neighboring motion vectors so that it can produce a minimum bitrate in motion vector difference coding. Then the prediction error, or motion vector difference (MVD), and the mode information (MODE) for determining the predicted motion vector at a decoder are coded and transmitted in order. Sending bits for MVD ahead of bits for the MODE, the scheme can minimize the bit amount for the MODE by taking advantage of the fact that the minimum bitrate predictor is used for motion vector prediction. Adaptively combining this minimum bitrate prediction scheme with the conventional model-based prediction scheme, more efficient motion vector coding can be achieved. The proposed scheme improves coding efficiency noticeably for various video sequences.

Index Terms—Low bitrate video coding, minimum bitrate prediction, motion vector coding.

I. INTRODUCTION

Various video coding schemes have been developed to reduce bitrates without seriously degrading image quality. However, these activities have been focused on reducing bits for texture coding because bits spent on other information such as motion vectors and syntax were considered negligible. In video coding at very low bitrates, however, the bit portion for motion vectors and syntax is not negligible anymore. Especially, the number of motion vector bits plays an important role in the overall coding performance in very low bitrate coding [1], [2]. There are two approaches to reduce the amount of motion vector bits. The first approach focuses on finding an appropriate motion vector estimation scheme to reduce the overall motion vector entropy without noticeably degrading motion compensation performance. This approach includes the motion field smoothing algorithm [3] and the vector quantization algorithm of motion vectors [4]. The second approach, which will be investigated in this work, involves finding a more efficient coding scheme for already estimated motion vectors. Since these algorithms are based on differential pulse code modulation (DPCM) [5]–[7], the proper selection of a predictor based on motion vector correlation is an essential element of coding efficiency.

In the conventional block-based coding scheme, the motion vector of each block is predicted as one of its neighboring motion vectors, and then the difference between the true (or current) and predicted motion vector is compressed using a variable-length coder. Since motion vector coding is lossless, the coding performance mainly de-

pends on the predictor performance. Currently, the median predictor is widely used in state-of-the-art low bitrate coding schemes such as H.263 [8]. In many sequences having fast or complex motions, however, the median predictor is not good enough for minimizing the number of bits spent on motion vector coding.

In this correspondence, we propose a new motion vector coding scheme based on minimum bitrate prediction instead of a model-based prediction scheme such as median prediction. For minimum bitrate prediction, the current block motion vector is predicted as one of the three causal neighboring motion vectors so that it can produce a minimum bitrate in motion vector difference coding. And then mode information for determining the predicted motion vector in a decoder and the prediction error are coded and transmitted. Although minimum bitrate prediction gives the best results in terms of prediction error, the required mode information prevents it from being widely used. In order to reduce its intrinsic burden and make it more efficient than median prediction, we propose a new bitstream syntax of variable-length mode information. In addition, adaptively applying this new minimum bitrate prediction and the conventional model-based median prediction guarantees coding performance improvement over various video features. Experimental results show that the proposed scheme considerably reduces the number of bits for motion vectors, especially in complex sequences where model-based prediction does not work well (savings up to about 10% in bits for motion vector coding, in maximum).

II. PROPOSED ALGORITHM

A. Minimum Bitrate Prediction

For motion vector coding, the current motion vector MV_c is predicted as one of three neighboring motion vectors MV_1 , MV_2 , and MV_3 (see Fig. 1). To predict the current motion vector, we examine every neighboring motion vector candidate (MV_1 , MV_2 , and MV_3) and select the one that produces the minimum bitrate in motion vector difference coding. Then we encode and transmit the difference MVD_{mbp} between the true motion vector MV_c and the predicted motion vector. Since the number of prediction candidates is three at most and the bitrate for motion vector difference coding can be estimated easily, the computational burden to select the motion vector producing the minimum bitrate in the encoder is not significant. However, to determine which neighboring motion vector is selected as the predicted one, the mode information (MODE) has to be prepared and sent to the decoder.

B. Syntax for MVD Code Bits and MODE Bits

Fig. 2 shows the proposed bitstream syntax for MVD_{mbp} and MODE bits. In both cases, x - and y -components are treated separately. It should be noted that MVD_{mbp} is ahead of MODE because the number of MODE bits can be reduced by utilizing the MVD_{mbp} information.

C. MODE Bits Assignment

Since the number of candidates is three, 2 bits are needed at most for MODE bits. In the syntax proposed above, however, MODE bits can be reduced by considering the fact that the MVD_{mbp} transmitted

Manuscript received January 6, 1998; revised December 14, 1998. This work was supported by LG Semicon Co. Ltd., Korea. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Robert Forchheimer.

The authors are with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Taejeon, Korea (e-mail: jbra@ee.kaist.ac.kr).

Publisher Item Identifier S 1057-7149(99)06103-5.

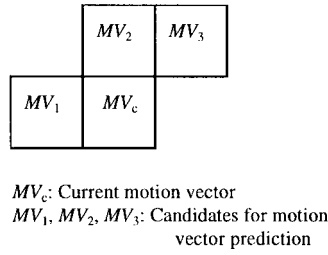


Fig. 1. Motion vector coding scheme.

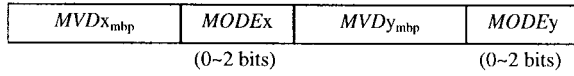


Fig. 2. Bitstream syntax for motion vector coding.

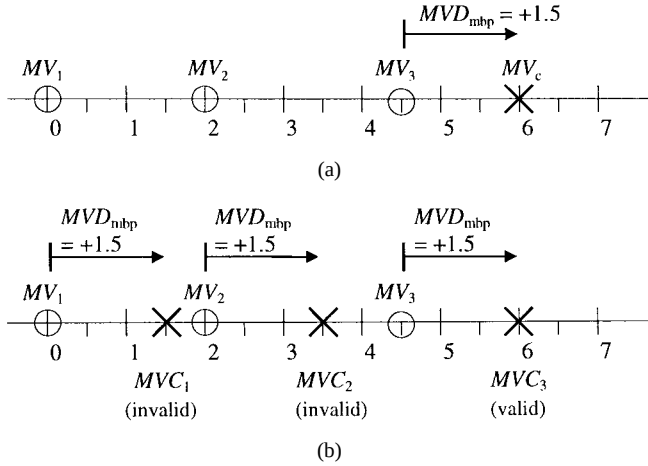


Fig. 3. MODE bit assigning example. (a) Encoder side. (b) Decoder side. In this case, no bit is needed for the MODE.

to a decoder is obtained on the basis of minimum bitrate prediction. Depending on the values of MV_1, MV_2, MV_3 , and MVD_{mbp} , the required number of bits for MODE varies from zero to two. How to assign the minimum number of MODE bits can be demonstrated by three different examples shown in Figs. 3–5, where zero, one, or two MODE bits are to be sent.

In Fig. 3, neighboring motion vectors $MV_1 = 0, MV_2 = 2.0$, and $MV_3 = 4.5$, and the current motion vector $MV_c = 6.0$. Then, MVD_{mbp} of $+1.5$ is coded in the encoder side and transmitted to a decoder. At the decoder side, three candidates for MV_c , MVC_1, MVC_2 , and MVC_3 , can exist on the basis of the MVD_{mbp} . However, MVC_1 cannot be the true motion vector because MVD_{mbp} for MVC_1 is not $+1.5$ but -0.5 . Similarly, MVC_2 cannot be the true motion vector because MVD_{mbp} for MVC_2 is not $+1.5$ but -1.0 . Therefore, the true motion vector is MVC_3 . It is noted in this case that no MODE bit is needed at the decoder to find the true motion vector.

For the second example, Fig. 4 shows that $MV_1 = 0, MV_2 = 3.0, MV_3 = 4.5$, and $MV_c = 1.0$. In this case, MVD_{mbp} of $+1.0$ is coded in the encoder and transmitted first. At the decoder side, candidates for MV_c are also MVC_1, MVC_2 , and MVC_3 . But MVC_2 cannot be the true motion vector because MVD_{mbp} for MVC_2 is not $+1.0$ but -0.5 . Therefore, the true motion vector is either MVC_1 or MVC_3 . In this case, 1 bit is needed at the decoder to find the true motion vector.

Finally, Fig. 5 demonstrates the example of $MV_1 = 0, MV_2 = 3.0, MV_3 = 4.5$, and $MV = 3.5$. At the encoder side, MVD_{mbp}

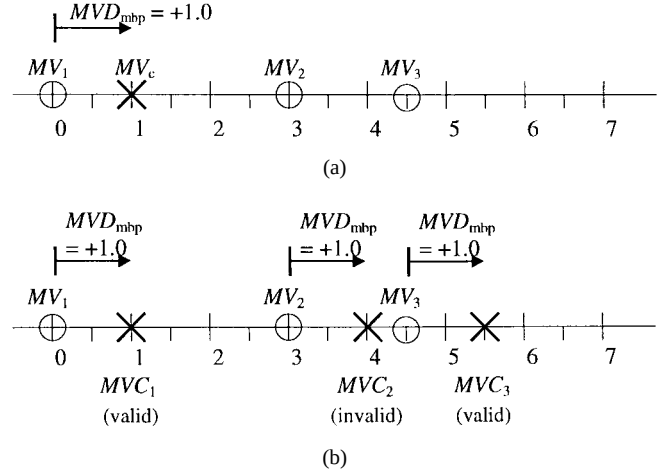
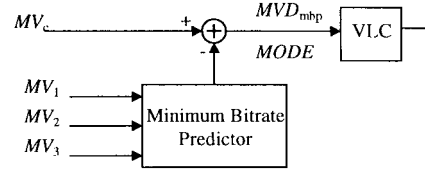


Fig. 4. MODE bit assigning example. (a) Encoder side. (b) Decoder side. In this case, one bit is needed for the MODE.

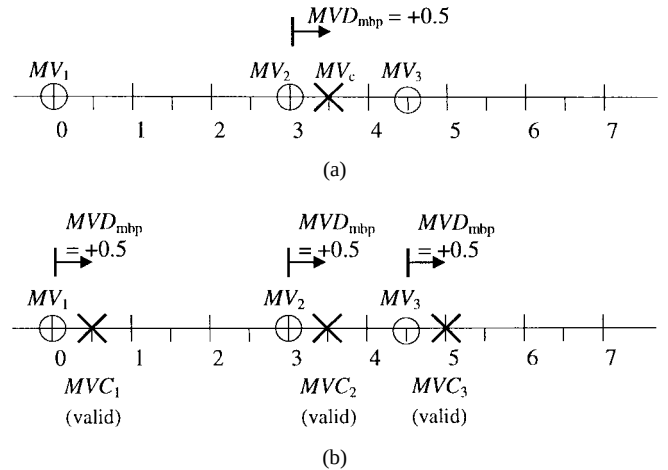


Fig. 5. MODE bit assigning example. (a) Encoder side. (b) Decoder side. In this case, one or two MODE bits are needed for the MODE.

TABLE I
SYNTAX FOR MODE BITS

No. of valid MVC_i	True motion vector among valid MVC_i	No. of bits	Code
1	Unique	0	x
2	Maximum one	1	"0"
	Minimum one	1	"1"
3	Median one	1	"0"
	Maximum one	2	"10"
	Minimum one	2	"11"

TABLE II
MOTION VECTOR CODING RESULTS USING MINIMUM BITRATE PREDICITON (MBP) WITH AND WITHOUT MODE BIT REDUCTION

Bit-rates	Sequences	Res.	FR	Bits for motion vector coding (300 frames)				
				VM	MBP without <i>MODE</i> bit reduction		MBP with <i>MODE</i> bit reduction	
					Total	<i>MODE</i> only	Total	<i>MODE</i> only
24 kbps	Hall monitor	QCIF	10	18369	19589	5242	18590	4243
	Container ship	QCIF	10	12709	13162	2022	13091	1951
	Mother&daughter	QCIF	10	39956	42136	12829	39701	10394
	Silent voice	QCIF	10	72543	71589	20472	66764	15647
48 kbps	Foreman	QCIF	10	138584	135887	38172	127857	30142
	Coast guard	QCIF	10	83430	82742	24217	78773	20248
	News	CIF	7.5	142087	136723	39684	128546	31507
112 kbps	Foreman	CIF	15	673168	645013	85637	606819	77399
	Coast guard	CIF	15	516648	528274	170246	496718	138690
	News	CIF	15	212377	210819	62659	197979	49819

TABLE III
EXPERIMENTAL RESULTS OF MOTION VECTOR CODING AND TOTAL IMAGE SEQUENCE CODING USING THE VM PREDICTION SCHEME AND THE PROPOSED SCHEME WITH THR OF 1.0, RESPECTIVELY

Bit-rates	Sequences	Res.	FR	Bits for motion vector coding (300 frames)			Total bits for image sequence coding (300 frames)		
				VM	Proposed	Savings [%]	VM	Proposed	Savings [%]
24 kbps	Hall monitor	QCIF	10	18369	18281	0.48	245064	244896	0.07
	Container ship	QCIF	10	12709	12703	0.05	265176	265168	0.00
	Mother&daughter	QCIF	10	39956	39308	1.62	234960	234336	0.27
	Silent voice	QCIF	10	72543	66606	8.18	245304	239432	2.39
48 kbps	Foreman	QCIF	10	138584	127504	8.00	481104	469984	2.31
	Coast guard	QCIF	10	83430	77697	6.87	510328	504624	1.12
	News	CIF	7.5	142087	128327	9.68	461088	447344	2.98
112 kbps	Foreman	CIF	15	673168	606504	9.90	1233464	1166816	5.40
	Coast guard	CIF	15	516648	496534	3.89	1102928	1082816	1.82
	News	CIF	15	212377	196522	7.47	1160648	1144768	1.37

of $+0.5$ is coded and transmitted first. In this case, MVC_1 , MVC_2 , and MVC_3 can be a candidate for MV_c at the decoder, because MVD_{mbp} for all MVC_i , $i = 0, 1, 2$, have the same value of $+0.5$. Therefore, 1 or 2 bits are needed for mode information.

Using the procedure adopted in the three examples above, we can assign MODE bits for any set of three neighboring and current motion vectors. The syntax for MODE bits is given in Table I.

D. Adaptive Application of Minimum Bitrate Prediction

Minimum bitrate prediction has a considerable gain in motion vector coding compared to a model-based prediction scheme such as median prediction, when the prediction model does not coincide with the actual situation. But, in video sequences with very homogeneous motions, model-based prediction may provide a slightly better performance. Therefore, if we adaptively combine the minimum bitrate prediction scheme with the model-based one according to the distribution of neighboring motion vectors, we can expect improvement in coding efficiency.

As a measure to find a case when the median predictor works well, we use $\Delta = |MV_{\min} - MV_{\max}|$, where $MV_{\min}$ and $MV_{\max}$ are the minimum and maximum motion vectors among the three neighboring ones, respectively. Since Δ tends to represent a spreading degree of neighboring motion vectors, median prediction is more likely to succeed for $\Delta \leq THR$. Therefore, if $\Delta \leq THR$, we apply the median predictor, otherwise, the minimum bitrate predictor.

III. EXPERIMENTAL RESULTS

Simulation was performed on the basis of the MPEG-4 verification model (VM) 7.0 [6], which uses the model-based median predictor for motion vector coding. We substituted the motion vector prediction part with our proposed version but used the same variable length coder (VLC) table. In the simulation, the motion search range of -16.0 to 15.5 was used, and the advanced prediction mode was enabled. Various MPEG-4 test sequences that have different amounts of movement were used for proper evaluation of the proposed scheme.

To examine the MODE bit reduction effect in motion vector coding, we compared bits for motion vectors for three cases, i.e., median prediction in VM and minimum bitrate prediction with and without MODE bit reduction (see Table II). As shown in the table, the conventional minimum bitrate prediction without MODE bit reduction increases the amount of bits compared with median prediction in VM. However, the proposed MODE bit reduction scheme reduces the MODE bit amount (up to about 20%) so that minimum bitrate prediction becomes more efficient than median prediction. According to this result, MODE information is a critical factor in coding efficiency. However, bit amounts for MV coding are slightly larger in a couple of sequences containing mainly small and homogeneous motion vectors. This is because the median predictor is more efficient for homogeneous motions than the minimum prediction predictor which requires mode information. Hence, the proposed adaptive prediction is a good candidate for solving this problem.

Table III shows the experimental results from the adaptive motion vector prediction and compares them with those of the ordinary median prediction in VM. For this experiment, THR was heuristically determined and set to one. As shown in the table, by using the proposed adaptive scheme, the amount of bits for motion vector coding is saved consistently for all the test sequences with various bit rates. And also, total bit savings in sequence coding demonstrate that the bit portion for motion vector information is big enough to be considered seriously in very low bitrate coding. The number of bits needed for motion vector coding is reduced especially in the video sequences with complex or random motion fields, such as in the foreman, silent voice, and news sequences. This is because, for complex or random motion fields, the proposed scheme performs better than the conventional scheme based on median prediction.

IV. CONCLUSIONS

Even though minimum bitrate prediction in motion vector coding fully exploits motion vector correlation with neighboring blocks, model-based prediction has been preferred because it does not require extra MODE information indicating the selected neighboring motion vector unlike minimum bitrate prediction. In this work, we have proposed a novel scheme to efficiently utilize the minimum bitrate prediction scheme by minimizing mode information bits. By combining minimum bitrate prediction and median model-based prediction properly, we can noticeably reduce the bit amount for motion vectors. We have shown that this method is well suited to the further reduction of motion vector information in very low bitrate coding, especially for complex or random video sequences.

REFERENCES

- [1] A. Nicoulin, M. Mattavelli, and M. Kunt, "Motion field coding using a composite source model," in *Proc. Int. Workshop Coding Tech. Very Low Bit-Rate Video (VLBV94)*, U.K., Apr. 1994.
- [2] P. Guillotel and C. Chevance, "Comparison of motion vector coding techniques," in *Proc. SPIE Visual Communications and Image Processing*, Sept. 1994, pp. 1594–1604.
- [3] C. Stiller, "Motion-estimation for coding of moving video at 8 kbit/s with Gibbs modeled vectorfield smoothing," in *Proc. SPIE Visual Communications and Image Processing*, Oct. 1990, pp. 468–476.
- [4] Y. Y. Lee and J. W. Woods, "Motion vector quantization for video coding," *IEEE Trans. Image Processing*, vol. 4, pp. 378–382, Mar. 1995.
- [5] G. Y. Yu and C.-T. Chen, "Two-dimensional motion vector coding for low bitrate videophone applications," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1995, pp. 414–417.
- [6] K. Illgner and F. Muller, "Hierarchical coding of motion vector fields," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1995, pp. 566–569.
- [7] J. Yeh, M. Vetterli, and M. Khansari, "Motion compensation of motion vectors," in *Proc. IEEE Int. Conf. Image Processing*, Oct. 1995, pp. 574–577.
- [8] ITU-T Study Group 15, "ITU-T recommend. H.263," May 1996.
- [9] ISO/IEC JTC1/SC 29/WG 11/MPEG97/N1642, "MPEG-4 video verification model version 7.0," Apr. 1997.

Minimum Error Rate Training for PHMM-Based Text Recognition

Chinching Yen, Shyh-shiaw Kuo, and Chin-Hui Lee

Abstract—In this work, discriminative training is studied to improve the performance of our pseudo two-dimensional (2-D) hidden Markov model (PHMM) based text recognition system. The aim of this discriminative training is to adjust model parameters to directly minimize the classification error rate. Experimental results have shown great reduction in recognition error rate even for PHMM's already well-trained using conventional maximum likelihood (ML) approaches.

Index Terms—Discriminative training, document image processing, hidden Markov models, image recognition, learning system, optical character recognition, pattern recognition, stochastic models.

I. INTRODUCTION

One of the approaches to the text image recognition problem is to use statistical pattern recognition method. Statistical model parameters are estimated over a training set by optimizing a certain criterion. For example, maximum likelihood (ML) scheme has been one of the most popular optimization criteria for model estimation in statistical pattern classification. The scheme attempts to find the optimal parameter set that maximizes the class-conditioned observation probability (the likelihood) over the training set. Due to the lack of the exact knowledge about the statistical properties of the observations, the model parameters obtained through this ML training do not necessarily achieve minimum classification error (MCE).

Recently, discriminative training (DT) has been proposed and studied in speech recognition [1]–[4]. Instead of adapting model parameters to achieve maximum overall likelihood score, the objective of this training is to minimize classification error through parameter estimation. Both correct and incorrect yet competitive models are combined as part of the optimization formulation, which in turn is directly linked to the optimization criterion—the classification error rate. Therefore, through this MCE training, parameter adjustments can directly contribute to minimizing classification errors.

In this correspondence, MCE training is developed for text recognition in a pseudo two-dimensional (2-D) hidden Markov model-based (PHMM-based) system. The "competing models" are found through our "duration-corrected N -best hypotheses search" [5], where duration information and correction is incorporated in the multiple hypotheses search. The experimental results have shown that MCE training approach greatly improves the performance of the original ML-trained models, in terms of reducing classification errors. For details about PHMM-based system, or the duration-corrected N -best hypothesis search, please refer to [5], [6]. In this work, we will be using Λ to represent model set, λ_i for individual models. For PHMM, each λ_i contains, but not limited to, probability information about

Manuscript received January 23, 1997; revised May 5, 1998. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Dmitry B. Goldgof.

C. Yen is with Bell Laboratories, Lucent Technologies, Holmdel, NJ 07733 USA (e-mail: rcyen@lucent.com).

S. Kuo is with AT&T Laboratories—Research, Florham Park, NJ 07932 USA (e-mail: skuo@att.com).

C.-H. Lee is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: chl@research.bell-labs.com).

Publisher Item Identifier S 1057-7149(99)06105-9.