Title

Classification based fast mode decision for stereo video coding

Authors

Yu, T
Zhang, Y
Cosman, P C

Publication Date

2013-09-01

Peer reviewed

CLASSIFICATION BASED FAST MODE DECISION FOR STEREO VIDEO CODING

Ting Yu[1], *Yuan Zhang*[1], *Pamela C. Cosman*[2]

[1] Communication University of China, Beijing, 100024, China
[2] University of California, San Diego, CA, 92093-0407, USA

ABSTRACT

We propose a classification based fast mode decision scheme in stereo video coding. By treating mode decision as a classification problem, our scheme employs a decision tree classifier to separate out SKIP mode, which is the major and most computationally efficient mode in stereo video coding. Thus, we can pre-decide whether the current macroblock is coded as SKIP mode without going through the exhaustive mode decision process. Experimental results show that this scheme provides 30~75% of time saving over a wide range of quantization parameter values.

Keywords— *stereo video coding; fast mode decision; CART classifier*

1. INTRODUCTION

Three-dimensional (3D) video services, including 3DTV, FTV, immersive teleconferencing and 3D-surveillance, are expected to become more popular in multimedia industry. Various 3D video formats have been designed to support such services. These 3D formats include video-only formats (such as stereo and multiview video) and depth-enhanced formats (such as video plus depth or multiview video plus depth). In all these formats, at least two video/depth sequences and possibly additional depth data have to be represented. The enormous data rate makes efficient compression essential for 3D video applications [1].

As the most common 3D video format, stereo video can provide users with sense of depth perception by showing a different frame to each eye simultaneously. The prediction structure used in our work of stereo video coding is shown in Figure 1. The left-view sequence is encoded based on the conventional motion compensation architecture, while the right-view sequence is predicted from the previous reconstruction frame by motion estimation, as well as the corresponding decoded left-view by disparity estimation (DE). In this way, the stereo video coding exploits both temporal and inter-view correlation. For each macroblock (MB), the best motion vector, disparity vector, and coding mode are decided based on rate-distortion optimization (RDO), which evaluates every possible mode combination in the spatial, temporal, and disparity domains. This significantly increases the overall computational complexity. Thus, developing a fast mode decision scheme to reduce the computational complexity while maintaining almost the same R-D performance has become the main challenge in real-time 3D applications.

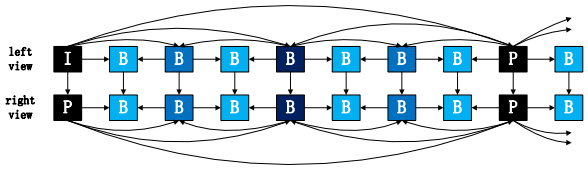

Figure 1. Stereo coding, with temporal and interview prediction

Many fast mode decision schemes for multi-view/stereo video coding have been proposed [2-8]. An adaptive early SKIP mode decision method exploits mode correlation in the 3D-neighborhood, variance, and RD properties and employs adaptive thresholds based on quantization parameters (QPs) [2]. In [3], the speed-up is achieved by avoiding checking the remaining modes once the RD cost of SKIP mode is below an adaptive threshold derived from the mode correlation. In [4], a mode complexity parameter was derived from the mode context of local MBs in the previously coded view, and the inter mode decision path was selected according to the mode complexity parameter. In [5], a fast disparity estimation and motion estimation algorithm based on motion homogeneity is proposed. An object-based mode decision method is proposed in [6], which encodes the foreground region using disparity estimation and the background region using motion estimation, while MBs at the boundaries are coded with the exhaustive RDO. A content-aware prediction algorithm with inter-view mode decision is proposed in [7]. By utilizing disparity estimation to find corresponding blocks between different views, the coding information (such as rate-distortion cost, coding modes and motion vectors) can be effectively shared and reused from the coded neighboring view. In [8], a fast inter mode decision method based on textural segmentation and correlations is presented.

Since mode decision is essentially a classification problem, in this paper we propose a fast mode decision scheme for stereo video coding using a CART decision tree. CART is short for Classification And Regression

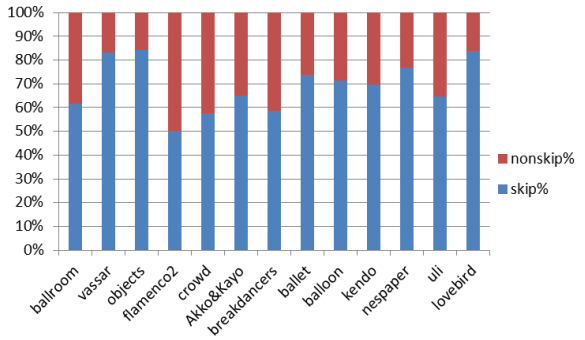

Figure 2. SKIP mode distribution of non-anchor pictures in right view (QP=28)

Tree, which is a well-developed method widely used in solving classification problems and is also easy to implement in practice [9][10]. The scheme starts with extracting computationally efficient features in the coding process. By growing a CART decision tree, we are able to select these features according to their usefulness in discerning mode classes. Computational complexity is reduced since only the modes belonging to the selected class need to be evaluated.

The rest of this paper is organized as follows. Section 2 describes the proposed method, including the feature selection and the subsequent classifier building. Experimental results are presented in Section 3. Finally, Section 4 concludes this paper.

2. PROPOSED ALGORITHM

In this section, the selection of the MB classes and the definition of class decision metrics (features) are examined. The class decision metrics should be defined based on their computational complexity and accuracy in predicting MB classes. Finally, CART results are presented.

2.1 Selection of MB classes

Figure 2 shows that the SKIP mode takes up over 50% of all modes in all the sequences. The percentage of SKIP mode depends on the slice type, since I slices have no SKIP mode while B slices have the most and P slices lie in between. In the hierarchical B picture (HBP) prediction structure, anchor and non-anchor pictures present different distributions of coding modes due to their different slice types. Also, the percentage of SKIP decisions increases with larger QP. This can be explained by the RD cost function. When QP is large, λ_{MODE} is accordingly large and the bit rate R dominates the RD cost. Thus, the SKIP modes are most likely chosen. For different video sequences, the SKIP mode distribution varies with the video content. Usually, the background and motionless or low motion areas tend to be predicted in SKIP mode. If we can pre-determine whether the MB is coded as SKIP mode or not, computation in mode decision can be saved. Therefore, we separate the MB classes into SKIP and non-SKIP.

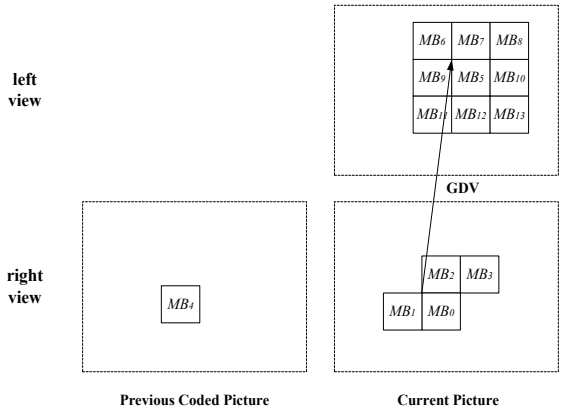

Figure 3. The neighboring MBs of the current macroblock

2.2 Study and selection of features

In this section, we describe the features that will be useful in the prediction of the mode decision. We consider the spatial, temporal and inter-view correlation between the current MB and the neighboring MBs. The coding mode, motion and texture characteristic of the neighboring MB set are specifically considered.

First, we should determine the neighboring MB set. As shown in Figure 3, the neighboring MBs consist of the spatially and temporally nearby MBs in the current and neighboring view [3]. The corresponding MB in the neighboring view (i.e., MB_5) is located based on the global disparity vector (GDV) between the current view and the neighboring view as introduced in [11].

We evaluate the coding mode information of the neighboring MBs to get the mode correlation. Two mode-related features, i.e. percentage of SKIP MBs in the neighboring MBs and mode complexity are considered. We modify the mode complexity in [4] as follows:

$$mode\_complexity = \frac{\sum_{i=1}^{N} W_i \cdot mc_i}{\sum_{i=1}^{N} W_i} \quad (1)$$

where W_i is the weight factor assigned based on the position of neighboring MBs. Compared with MBs in the diagonal direction, the MBs in the horizontal and vertical direction are more important in the mode decision. Accordingly, the weight factors are defined as shown in Table 1. While mc_i is the mode-weight factor assigned based on the complexity of each mode. The *mc* of various modes are shown in Table 2. We get these factors from [3] and [4], and modify the factors empirically to fit our situation.

Table 1: weight factor of neighboring MBs based on position

MB's Index i	1,2,4,5	3,7,9,10,12	6,8,11,13
W_i	1.30	0.96	0.75

Table 2: mode-weight factor of various modes

Modetype	SKIP	16×16	16×8	8×16	8×8	Intra
mc	0.5	1	2	2	3	4

Features related to motion activity are then evaluated. We first evaluate the motion vector (MV) strength in the neighboring MBs, including the average MV, the max

MV and the minimum MV. Then, we consider motion homogeneity defined in [5] and adapt it to our situation. The motion homogeneities in horizontal and vertical directions are respectively defined as,

$$MD_x = \frac{1}{64}\sum_{(i,j)\in Z}\left|mvx_{i,j} - \frac{1}{64}\sum_{(u,v)\in Z} mvx_{u,v}\right| \quad (2)$$

$$MD_y = \frac{1}{64}\sum_{(i,j)\in Z}\left|mvy_{i,j} - \frac{1}{64}\sum_{(u,v)\in Z} mvy_{u,v}\right| \quad (3)$$

where Z represents the 4×4 blocks covered by an MB and its closest neighboring MBs (MB_1,MB_2,MB_3,*and* MB_0). The motion homogeneity of the current MB is defined as,

$$MD = (MD_x + MD_y)/2 \quad (4)$$

Since only after ME is performed, motion vector information of an MB is available. Besides, the motion vectors of an MB in one view are strongly related to those of the corresponding MB in the previous coded views [5]. Thus when considering MD to predict the motion homogeneity of the current MB, we substitute motion vector of the uncoded current macroblock MB_0 with that of corresponding macroblock MB_5.

Textural features are also examined. We evaluate the variance of the current MB before encoding as [2]:

$$Var_{MB} = \frac{1}{256}\sum_{i=1}^{256}\left(P_i - P_{AVG}\right)^2 \quad (5)$$

$$P_{AVG} = \left(\sum_{i=1}^{256} P_i + 128\right) >> 8 \quad (6)$$

here P_i is the luminance value of ith pixel in an MB. Table 3 summarizes all features used in coding mode classification.

Table 3: Description of features in mode classification

Features	Description
skip_num	Num. of SKIP mode of neighbor MBs
mode_complexity	Mode complexity of neighbor MBs
AVG_MV	Average MV of neighbor MBs
MAX_MV	Maximum MV of neighbor MBs
MIN_MV	Minimum MV of neighbor MBs
MD	Predictive motion homogeneity of the current MB
variance	Variance of current MB's pixels
curr_QP	QP value of current MB

2.2 CART results

We use CART decision tree for classification and further mode predication [12].

In the training stage, we use seven different 1024×768 sequences, which are Breakdancers, Ballet, Balloons, Kendo, Newspaper, Uli and Lovebird. Those sequences are encoded by H.264 reference software JM18.0, using the Stereo High Profile. The context-adaptive binary arithmetic coding (CABAC) is used as entropy coder, and the variable prediction size and the loop filter are turned on. Hierarchy is used and B reference is set as 7. The search range of ME and DE is ±64. The QP is set at 24, 28, 32 and 36. For each MB, the encoder does an RDO-based exhaustive search of all coding modes to get the best mode.

The training data set includes the values of the features in Table 3 and the best coding mode of each MB, marked as SKIP or non-SKIP. We use SPSS Clementine v11.1 software [13] to process data set to build the CART model. For each QP, we get one classifier. Each classifier is a decision tree; the classifier traverses a tree where the path at each node depends on a binary decision using one of the features in Table 3. During the formation of the tree, a node is split to minimize the probability of misclassification.

In the testing stage, we use six 640×480 sequences including Ballroom, Vassar, Objects2, Flamenco2, Crowd and Akko&Kayo. The process and experiment setting for obtaining the testing data set are the same as those for the training data set.

The performance of the four classifiers (for the four different QP values) is shown in Table 4. Entries under "TR" correspond to the performance during the training phase. Entries under "TEST" correspond to the validation phase, which classifies the testing data. "Misclassification" indicates the percentage of misclassification of non-SKIP to SKIP error, while "Accuracy" indicates the overall discriminant accuracy. We can see the CART decision tree classifiers in different QPs represent a compromise between misclassification and accuracy. Overall the classifiers have acceptable accuracy. When QP=36 the misclassification rate differs from the former ones, because the misclassification cost in QP=36 is set higher and thus the misclassification of non-skip to skip mode in TEST occurs less.

Table 4: Misclassifications and accuracy for each classifier, during training and testing

QP	Misclassification		Accuracy	
	TR	TEST	TR	TEST
24	19.33%	17.82%	88.07%	87.77%
28	24.45%	17.06%	88.47%	88.69%
32	30.88%	25.05%	88.83%	88.63%
36	26.69%	4.16%	87.35%	81.96%

Figures 4 and 5 shows CART decision tree classifiers for QP=24 and QP=32 with misclassification cost 1.0. We can modify the model by adjusting the misclassification cost. The misclassification of non-SKIP to SKIP may decrease the coding efficiency, while misclassifying SKIP to non-SKIP only increases the coding time. So, we should minimize the misclassification of non-SKIP to SKIP error under the premise of ensuring the overall discriminant accuracy. In Figure 4, the skip_num, curr_QP and MD are the most important features in the mode decision process. While in Figure 5, the decision tree contains 9 terminal nodes, and the initial split is based on skip_num. In this case, features such as skip_num, curr_QP, mode_complexity and variance are more important in mode decision.

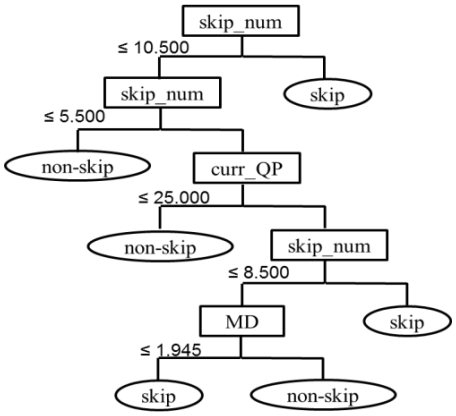

Figure 4. CART decision tree in QP 24 with cost 1.0

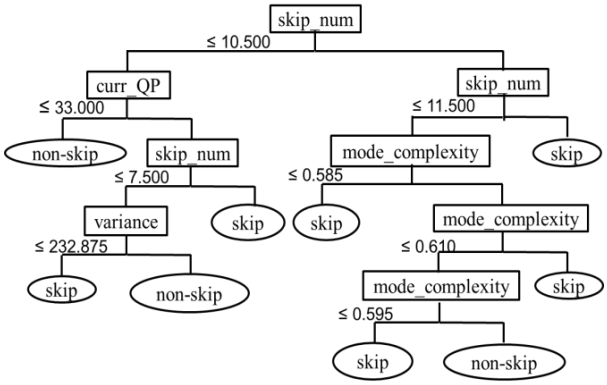

Figure 5. CART decision tree in QP 32 with cost 1.0

3. EXPERIMENTAL RESULTS

The proposed algorithm was implemented in H.264 reference software JM18.0. The RDO-based exhaustive method was taken as a reference for comparison. Six 640×480 sequences including Ballroom, Vassar, Objects2, Flamenco2, Crowd and Akko&Kayo are used.

Comparing the proposed algorithm to the reference, we use ΔPSNR and ΔBitrate to indicate the gains in video quality measured at the encoder and the percentage increase of bit rate, respectively. As shown in Table 5, the proposed algorithm achieves similar R-D performance to exhaustive mode decision approach while reducing the coding time significantly. The time savings ranges from 30% to 75%.

The time savings from the proposed scheme is mainly achieved from the pre-determination of the skip/non-skip path. It depends on the percentage of skip-coded blocks. For example, the time savings on Flamenco2 is less than others, because it has fewer skip-coded blocks. For those video sequences which have complex texture information and fast moving objects, the misclassification may lead to lower coding efficiency.

For future work, fast inter mode selection should be considered to achieve better R-D cost prediction and yield greater time savings. The features can also be used to classify the inter mode partition and distinguish between DE and ME with more stages in the classifier. The method can also be further extended to MVC with more views, by using the JMVM reference software which performs GDV itself.

Table 5: Performance evaluation for the proposed scheme

QP	Sequence	ΔPSNR/dB	ΔBitrate/%	Time saving/%
24	Ballroom	0.118	-2.11	48.39
	Vassar	0.115	-8.60	67.08
	Objects2	-0.086	-6.86	71.29
	Flamenco2	0.136	-0.65	29.83
	Crowd	0.26	-7.12	44.13
	Akko&Kayo	-0.033	0.13	41.79
28	Ballroom	-0.053	-0.84	47.99
	Vassar	-0.249	0.79	72.42
	Objects2	-0.089	-1.92	73.22
	Flamenco2	-0.105	-1.00	36.52
	Crowd	0.247	1.86	44.15
	Akko&Kayo	-0.123	3.10	40.47
32	Ballroom	-0.765	-2.88	55.97
	Vassar	-0.196	-0.56	70.78
	Objects2	-0.048	-4.97	70.76
	Flamenco2	-0.545	-3.06	51.21
	Crowd	0.233	-3.28	53.64
	Akko&Kayo	-0.359	-3.17	57.47
36	Ballroom	-0.59	2.53	46.94
	Vassar	-0.116	-3.13	70.21
	Objects2	-0.066	-1.57	69.57
	Flamenco2	0.045	-2.62	53.66
	Crowd	0.228	-2.18	54.71
	Akko&Kayo	-0.351	-0.62	56.78

4. CONCLUSION

In this paper, we proposed a classification based fast mode decision algorithm, by using a CART decision tree model to analyze and obtain the useful features to build an efficient CART classifier. The time savings is achieved by using the CART classifier to pre-decide whether the mode is SKIP or not. The proposed algorithm exhibits considerable speedup for sequences without knowing any pre-encoding information of the current MB, especially sequences with large motionless areas.

5. ACKNOWLEDGMENT

This work was supported in part by National Natural Science Foundation of China (61001177).

6. REFERENCES

[1] P. Merkle, K.Müller, and T. Wiegand,"3DVideo Coding - An overview of Present and Up coming Stan-dards" *In Visual Communications and Image Processing 2010.Pro c. of SPIE Vol. 7744*, 2010

[2] B. Zatt, M. Shafique, S. Bampi, J. Henkel, "An Adaptive Early Skip Mode Decision Scheme for Multiview Video Coding", *Picture Coding Symposium*, pp.42-45, 2010.

[3] H. Zeng, K.-K. Ma, C. Cai, "Mode-correlation-based early termination mode decision for multi-view video coding", *ICIP*, pp. 3405-3408 ,2010.

[4] L. Shen, Z Liu, P An, R Ma, and Z. Zhang, "Low-Complexity Mode Decision for MVC", *Circuits and Systems for Video Technology*, pp.837-843, June 2011.

[5] L. Shen, Z. Liu, S. X. Liu, Z. Y. Zhang, and P. An, "Selective Disparity Estimation and Variable Size Motion Estimation Based onMotion Homogeneity for Multi-view

Coding," *IEEE Transactions on Broadcasting*, vol. 55, no. 4, pp. 761-766, 2009.

[6] S.-Y. Lee et al., "An Object-based Mode Decision Algorithm for Multi-view Video Coding", *ISM*, pp.74-81, 2008.

[7] L. F. Ding, P. K. Tsung, S. Y. Chen, W. Y. Chen, and L. G. Chen,"Content-aware prediction algorithm with inter-view mode decision for multiview video coding," *IEEE Trans. Multimedia*, vol. 10, no. 8, pp. 1553-1564, Dec. 2008.

[8] W Zhu, X Tian, F Zhou, and Y. Chen, "Fast inter mode decision based on textural segmentation and correlations for multiview video coding", *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 3, August 2010.

[9] L. Breiman, J. H. Friedman, R. A. Olshen,and C. J. Stone, "Classification and Regression Trees", *Chapman&Hall*, 1984.

[10] C. H. Lampert, "Machine Learning for Video Compression: Macroblock Mode Decision," *18th InternationalConference on Pattern Recognition (ICPR'06)*, 2006.

[11] H.-S. Koo, Y.-J. Jeon, and B.-M Jeon, "MVC Motion Skip Mode", *ISO/IEC JTC1/SC29/WG11 and ITU-T Q6/SG16, Doc. JVT-W081*, Apr. 2007.

[12] S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. A. Vaishampayan,"Modeling packet-loss visibility in MPEG-2 video", *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341-355, 2006.

[13] http://spss-clementine.software.informer.com/11.0/