Fast H.264 Encoding Based on Statistical Learning

Chen-Kuo Chiang and Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{ckchiang,lai}@cs.nthu.edu.tw

Abstract. In this paper, we propose an efficient video coding system that applies statistical learning methods to reduce the computational cost in H.264 encoder. The proposed method can be applied to many coding components in H.264, like intermode decision, multi-reference motion estimation, intra-mode prediction. First, representative features are extracted from video to build the learning models. Then, an off-line pre-classification approach is used to determine the best results from the extracted features, thus a significant amount of computation is reduced based on the classification strategy. The proposed statistical learning based approach is applied to the aforementioned three main components and a novel framework of learning based H.264 encoder is proposed to speed up the computation. Experimental results show that the motion estimation (ME) time of the proposed system is significantly speed up with twelve times faster than the H.264 encoder with a conventional fast ME algorithm, and the total encoding time of the proposed encoder is greatly reduced with about four times faster than the fast encoder EPZS in the H.264 reference code with negligible video quality degradation.

Keywords: Motion estimation, multiple-reference motion estimation, intermode decision, intra prediction, H.264, statistical learning, video coding.

1 Introduction

Due to the strong demand of storing and transmitting enormous amounts of video data, video compression has been a very important and practical problem in recent years. The H.264/AVC standard is the latest video coding standard developed by the ITU-T VCEG and MPEG. It provides good video quality at substantially lower bit rates than previous standards. In addition, it is designed to be flexible for a wide variety of applications, such as low and high bit rates, low and high resolution video, DVD storage, broadcast and multimedia telephony systems. To achieve such goal, several new coding tools are introduced into the H.264 standard, such as variable-block-size motion compensation, multi-reference motion estimation, directional intra prediction, in-loop deblocking filter and content-adaptive entropy coding, etc. However, high computational overhead comes along with these components as well. Thus, how to reduce the computational complexity in video coding while maintaining good video quality becomes an emergent goal for H.264 coding system.

Variable-block-size motion compensation is one of the key features in H.264 video coding. There are seven kinds of block sizes, 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and

G. Qiu et al. (Eds.): PCM 2010, Part II, LNCS 6298, pp. 179–189, 2010.

4x4. A 16x16 macroblock (MB) can be partitioned to 16x8, 8x16 or 8x8 sub-MBs. A sub-MB (8x8) can be further partitioned to 8x4, 4x8 or 4x4 blocks. Although it further reduces the coding bitrate and data redundancy in motion estimation, the computational complexity also increases significantly. Several methods have been proposed for fast intermode decision in H.264 encoding recently. One category is a semi-statistical learning approach that decides possible partition modes based on statistical analysis of various characteristics from a collection of training data. Kuo et al. [4] analyzed the likelihood and the correlation of motion fields for a suitable block mode selection. Zhan et al. [5] removed low-probability modes according to the correlative characteristics in MB mode selection and the statistical characteristics of sub-MB mode. Recently, Ma et al. [6] utilized the conditional motion cost to learn several thresholds. Only a subset of intermodes is selected as candidates for ME.

Multiple-reference-frame motion compensation is another useful component in H.264 coding. Using multiple reference frames can fully exploit temporal correlation in video sequences to achieve high video coding quality, especially under the conditions of object occlusion and non-rigid object transformation. Recently, many algorithms have been proposed for the multi-reference motion estimation problem. They observe that not every reference frame is useful for motion estimation. Thus, it turns out to be a reference frame selection problem to choose effective number of reference frames. The semi-statistical learning approach decides an appropriate number of reference frames based on statistical data analysis. Wu and Xiao [7] employed the statistical distribution of the reference frames along with some rules to determine the optimal reference frame number.

To improve the efficiency for intra prediction, Sim and Kim [8] presented an efficient mode decision algorithm based on the conditional probability of the best mode with respect to the best modes of the adjacent blocks. Joshi et al. [9] replaced the complex mode-decision calculations by a classifier which has been trained specifically to minimize the reduction in RD performance.

In this paper, we propose a general statistical learning framework to reduce the computational cost in H.264 encoder. The general approach can be easily applied to many coding components in H.264. The problems are formulated as classification problems in our approach. First, representative features are chosen according to the feature analysis from a number of H.264 encoded video sequences. Then, these features are used to train the sub-classifiers for some partial classification problems. After the training is finished, these sub-classifiers are integrated to build a complete classifier. Last, an off-line pre-classification approach is employed to generate all possible combinations of the quantized features and pre-classify them with the learned classifiers. The results are stored as a lookup table. During the run-time encoding, features are extracted and quantized. The best results can be determined by the learning table. Thus, the computation time of encoding can be significantly reduced. The proposed method is then applied to three components, intermode decision, multi-reference motion estimation and intra-mode prediction. We propose a new encoding system that integrates the new components and shows superior performance of the proposed schemes over previous methods through experiments. To the best of our knowledge, this is the first work that introduces a general machine learning approach and a learning-based system for efficient H.264 encoding. The rest of this paper is organized as follows. Section 2 reviews the conventional framework of H.264

encoder. Section 3 introduces the method to apply the general statistical learning approaches to a specific problem. The proposed statistical learning based H.264 encoder is presented in Section 4. Experimental results are shown in Section 5. Finally, Section 6 concludes this paper.

2 H.264 System Framework

In H.264 coding standards, it defines two dataflow paths, a forward path and a reconstruction path. A forward path includes an input frame F_n processed in units of a macroblock(MB). Each MB is encoded in intra or inter mode. For Intra prediction, an MB is predicted from samples in the current slice that have previously encoded, decoded and reconstructed. For Inter prediction, an MB is formed by motion-compensated prediction from one or more reference frames. Then, it is subtracted from the current block to produce a residual block that is transformed and quantized. A set of quantized transform coefficients are reordered and entropy-encoded. In a reconstruction path, the encoder decodes an MB to provide a reference for further prediction. The coefficients are scaled and inverse transformed to produce a difference block. Then it is added to create a reconstructed block. Fig. 1 shows the partial system flow of the H.264 encoder.

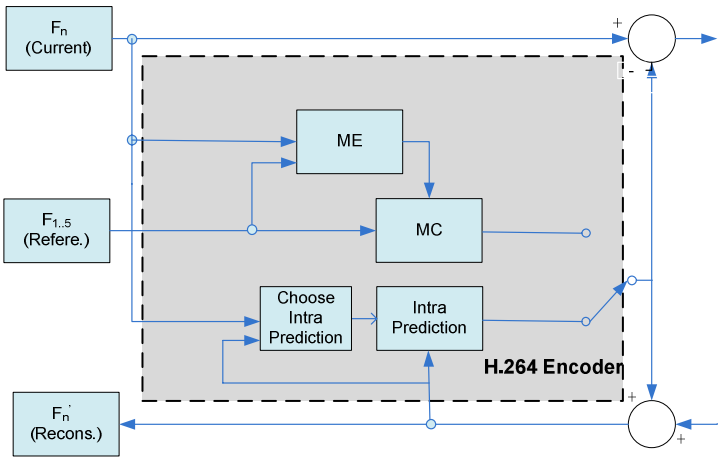

Fig. 1. The system framework of part of the H.264 encoder

3 Statistical Learning Approaches

In this section, we present a general approach to applying statistical learning methods to the H.264 encoding components. The proposed approach involves applying machine learning algorithms to develop a fast intermode decision algorithm [1], multi-reference frame number selection algorithm [2] and an efficient intra prediction algorithm [3].

3.1 Feature Selection

The first step of learning based approach is the feature selection. To choose effective features, several features which might be discriminative enough for the problem are chosen first. Take the problem of intermode decision for example, we examine the effectiveness of the feature Best Inter SAD, which means the best sum of absolute difference (SAD) of two MBs after applying block matching algorithm between inter-frames. The feature values are extracted during the encoding of a number of video sequences, and the corresponding intermode is also decided by the H.264 reference code. Then, the probability is calculated to show the relationship between the inter-mode and the selected features.

Fig. 2 shows the relationship between Best Inter SAD and the partition mode from 16x16 to 4x4. It indicates that lower SAD values correspond to higher probabilities of being a 16x16 or an 8x8 mode. In this case, such feature could be helpful.

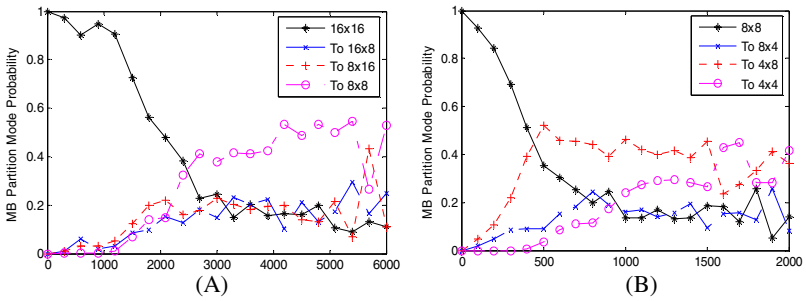

Fig. 2. The probabilities of different partition modes for 16x16 and 8x8 MBs for different *Best Inter SAD* values in the News sequence

The features used in our experiments are selected and described as follows:

Best Inter SAD. For an inter prediction MB, the ME procedure determines the best matching reference MB. The distortion measure used in the ME procedure in H.264 is the sum of absolute difference (SAD). SAD may indicate not only the accuracy of motion compensation but also the possibility of being a background MB. The lower the SAD value is, the higher probability the current MB contains still background. Thus, for a small Best Inter SAD value, it is very likely that this MB will not be split into sub-MBs.

Motion Vector Difference and Motion Vector Magnitude. Motion Vector Difference (MVD) is the sum of absolute value of the difference between the predicted MV and the motion vector after ME in horizontal and vertical directions. In the H.264 standard, the predicted MV is defined as the median of the MVs of the adjacent blocks in both x and y directions. MVD may represent the motion smoothness between current MB and adjacent MBs. If MVD is small, the current MB is more likely to be a background block. In this case, it is not necessary to partition this MB into sub-MBs.

Motion Vector Magnitude (MVM) is the sums of absolute values of all the motion vectors computed from the ME procedure in this MB. It indicates whether this MB is stationary or not. If the MB is stationary, it can be matched well by a large MB.

Best Intra SAD. Best Intra SAD is the minimal SAD value after the intra prediction of the current MB. An MB with a large SAD value after intra prediction usually contains object boundaries or complicated texture. Therefore, it tends to be partitioned into smaller sub-MBs.

Gradient Magnitude. The gradient magnitude of the current MB is defined as the summation of the gradient magnitudes of all pixels inside the MB obtained by applying the Sobel operator. The gradient magnitudes remain low in homogeneous regions. An MB with low gradient magnitude tends to be a background block, which is unlikely to be partitioned into sub-MBs.

Block Partition. In the process of motion estimation, MBs are partitioned into different block sizes from 16x16 to 4x4. The types of block partition are labeled from 1 to 7 for feature representation.

Neighboring Block Mode. The encoded block mode is closely related to its neighboring block modes. Based on the observation, we can predict most probable intra modes to achieve very efficient intra prediction.

3.2 Problem Formulation

The second step is problem formulation. In the learning based approach, we formulate the problem into a classification problem. For intermode decision problem, the seven partition types, from 16x16 to 4x4, could be considered as 7 different classes. For the multi-reference motion estimation problem, the number of reference frames might be defined as several classes. According to the analysis to Fig. 2 (A), two classes, C_1 and C_2, can be defined for mode 16x16 and non-16x16 by one binary classifier. Then, the binary classification results can be decided from the class conditional probabilities given the features. To be specific, the current MB is assigned to intermode 16x16 if the following inequality holds

$$P(C_1|feature_1, \ldots, feature_N) > P(C_2|feature_1, \ldots, feature_N) \quad (1)$$

Otherwise, class C_2 is assigned. To distinguish the rest mode from mode 16x8 to mode 4x4, other classes and classifiers can be designed by the users. In the end, all classifiers are integrated either in a parallel or cascade form to build a whole decision model.

3.3 Training and Off-Line Pre-classification

In the above, the decision rule is defined for the classification problem. However, it is difficult to model the joint probability of those features from limited training samples. The Support Vector Machine (SVM) [10] is used to solve the classification problem.

The training data is obtained by applying the H.264 reference code to several training sequences. Selected features are extracted when each MB is processed. The

intermode in the intermode decision problem decided by the reference code is regarded as the ground truth. Then, the training data is used to train different SVM classifiers based on our problem formulation. For the consideration of real-time encoding, it takes too much time for run-time classification in SVM. Thus, an off-line pre-classification strategy is exploited to minimize the computation time involved in the classification procedure. The idea is to generate all possible combinations of the quantized feature vectors and pre-classify them with the trained SVM classifier. To reduce the total number of possible combinations, a quantizer with adaptive step size is applied on the feature space. Features are quantized into several bins. The classification results are stored as a look-up table. During the encoding, the run-time features are extracted and quantized. By looking up the tables, the classification can be obtained easily and efficiently. Hence, the computation time can be significantly reduced by using this off-line pre-classification approach.

4 Proposed Statistical Learning Based H.264 Encoder

In this section, we propose the system framework for a statistical learning based H.264 encoder by integrating three major components, named *Intermode Classifier*, *MR Selector* and *Intramode Selector*, from our previous work, intermode decision algorithm [1], multi-reference motion estimation algorithm [2] and intra prediction algoritm [3], respectively. *Intermode Classifier* is used to decide those partition modes on which the ME should be applied. *MR Selector* determines the number of reference frame the current MB will use for variable block size motion compensation. An intramode is chosen by the *Intramode Selector* to perform intra prediction.

Feature selection and model training are accomplished as described in Section 3. All training processes are finished off-line. During the encoding, the trained models are loaded in the beginning. As depicted in Fig. 3, the encoding process first collects

Table 1. The encoding flow of the proposed statistical learning based H.264 encoder

The Procedures of the Learning Based H.264 Encoder	
Step 1	Collect features for intra mode decision.
Step 2	Select the best intra mode.
Step 3	Perform 16×16 ME using first reference frame.
Step 4	Collect features for the selection of the number of reference frames; collect features for intermode decision.
Step 5	Determine the required number of reference frames based on the learning model.
Step 6	Determine the intermodes based on the learning model.
Step 7	Finish all the required variable block size motion estimation and perform on the all selected reference frames.
Step 8	Select the best MB partition as the intermode.
Step 9	Select the best mode from intra and inter modes.
Step 10	Go to Step 1 and proceed to the next MB.

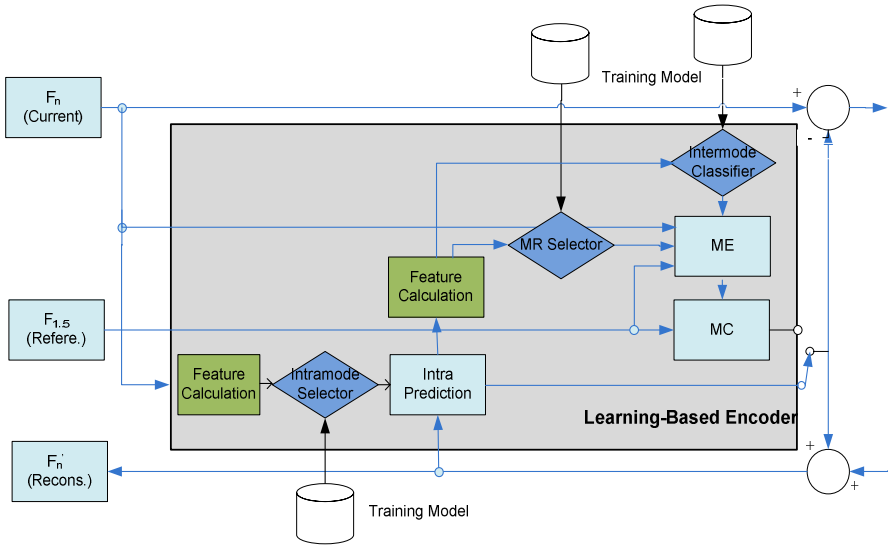

Fig. 3. The flowchart of the proposed H.264 encoding system by integrating three learning-based components, i.e. Intermode Classifier, MR Selector and Intramode Selector

the required features for intra prediction. Then, the best intramode can be decided by the training model and the run-time features. The main reason to move the procedure of intra prediction forward as the first step, which is different from the conventional H.264 encoder, is to collect features from intra prediction procedure for the next two components, *Intermode Classifier* and *MR Selector*. Next, ME is performed on the current MB for partitioning 16x16 on the first reference frame. Then, run-time features for *Intermode Classifier* and *MR Selector* are collected. These two classifiers will decide a subset of partition modes and the number of reference frames that the ME will be performed. Thus, unnecessary modes and reference frames are skipped by ME. Last, the encoder will choose the best mode from intermode or intramode based on the results from RDcost computation. The detailed steps are described as Table 1.

5 Experimental Results

We implement a statistical learning based H.264 encoding system based on the reference code JM11.0. The motion search range is set to 32 and the maximal number of reference frames is set to 5. The RD optimization and the CABAC entropy encoding are enabled. The transform 8x8 and the new intra 8x8 for luma component are off. The training data for fast intermode decision is obtained by applying the H.264 reference code to four video sequences; namely, News, Akiyo, Foreman and Coastguard. Three video sequences: News, Container and Coastguard videos are used for training in fast multi-reference motion estimation. For Intra prediction, Foreman, Container and Coastguard are used to extract features. All test sequences are in CIF format and tested on an Intel Core2 CPU 6320 at 1.86 GHz.

Table 2. The ME time of overall performance compared with Full Search for EPZS and the proposed method when QP is set to28 on CIF sequences with 300 frames

Sequences	ME Time (s)		
QP (28)	**FS**	**EPZS**	**Proposed**
HallMonitor	5868.69	160.40	16.71
M_D	5953.90	182.83	16.46
Stefan	5870.24	317.20	27.09
Akiyo	5850.68	146.94	14.85
News	5881.23	182.44	17.97
Coastguard	5854.72	341.68	19.86
Average	5879.91	221.92	18.82

Table 3. The total encoding time of applying the full search with EPZS and the proposed method when QP is set to28 to six CIF sequences with 300 frames

Sequences	Total Encoding Time (s)		
QP (28)	**FS**	**EPZS**	**Proposed**
HallMonitor	7962.87	2209.84	573.20
M_D	7778.16	1963.80	522.83
Stefan	8810.04	3241.64	710.56
Akiyo	7662.77	1942.42	511.82
News	7993.41	2349.74	566.41
Coastguard	8515.10	2953.35	646.45
Average	8120.39	2443.47	588.55

Table 4. PSNR and Bitrate results of applying the full search with EPZS and the proposed method to six CIF sequences with 300 frames when QP is set to 28

Sequences	PSNR decreased (dB)		Bitrate increased (%)	
(QP 28)	EPZS	Proposed	EPZS	Proposed
HallMonitor	0.01	0.04	0.00	7.09
M_D	0.01	0.09	0.00	9.10
Stefan	0.01	0.15	0.01	12.79
Akiyo	0.00	0.07	0.04	8.26
News	0.01	0.07	0.15	7.10
Coastguard	0.00	0.08	0.02	7.94
Average	0.007	0.083	0.037	8.71

We compare the proposed system with FS and EPZS on several testing video sequences. Table 2 shows the overall performance of ME time. It shows that the proposed encoding system is 324.25 times faster than the FS, while the EPZS is 29.48 times faster than FS in terms of ME time. In other words, the proposed system can achieve about 11.61 times faster than EPZS in ME time. Table 3 shows the overall performance of total encoding time. The proposed H.264 encoding system is 13.9

times faster than FS and 4.12 times faster than EPZS. Fig. 4 illustrates the speedup ratio based on the overall execution time. Table 4 shows the PSNR and bitrate about the proposed system. In average, the proposed system is 0.083 dB lower than FS while the bitrate is increased by 8.71%. It indicates that the proposed learning based H.264 encoder is efficient for video coding and effective in term of PSNR with slight bitrate increase. Fig. 5 depicts the RDcurves of FS, EPZS and the proposed method.

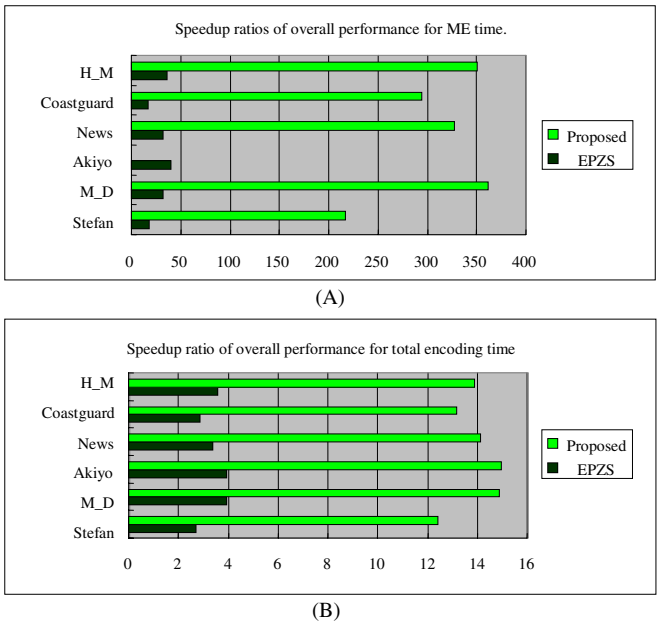

Fig. 4. Average speedup ratios of (A) ME time and (B) total encoding time of overall performance for the full search with EPZS and the proposed method when QP is set to 28 on six CIF sequences with 300 frames

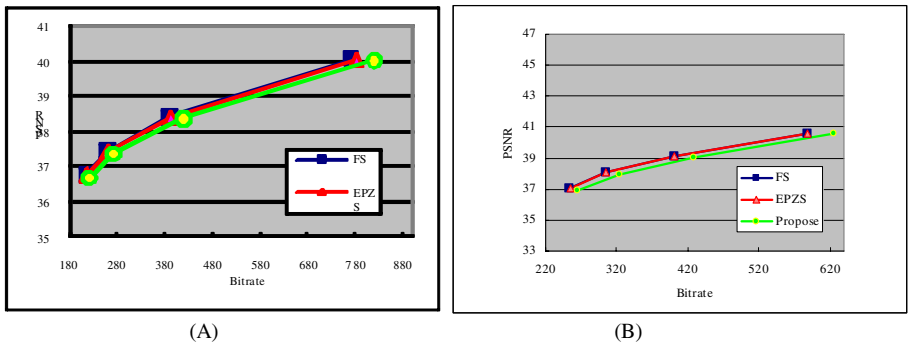

Fig. 5. RDcurve comparison of the EPZS, Full Search, and the proposed algorithm on (A) HallMonitor and (B) News sequence with 300 frames when QP is set to 24, 28, 32 and 36

6 Conclusion

In this paper, we present a statistical learning approach for efficient H.264 video encoding. The first step includes statistical feature analysis to find representative features. Then, the mode decision problem is formulated as a classification problem. To train the classifiers, features and ground-truth modes are collected by a number of H.264 encoded training video data. Learning models are trained by SVM. An off-line pre-classification approach is provided to speedup the classification procedure during the run-time. We apply the proposed algorithm to build three learning based classifiers for intermode decision, multi-reference motion estimation and intra-mode decision. With the above new components, we propose a novel fast statistical learning based H.264 encoding system.

To demonstrate the efficiency and effectiveness of the proposed system, experimental results are provided with comparisons to the existing methods. The overall performance of the entire system is improved about 12 times faster than EPZS in ME time and about 4 times faster than EPZS in total encoding time with negligible PSNR increase and slight bitrate increase.

To the best of our knowledge, this is the first work that introduces a general statistical learning approach for several H.264 coding components and provides a complete framework for H.264 video encoder. Experimental results show that the execution speed of our algorithm is significantly improved over the existing fast ME method while achieving slightly degraded compression quality in terms of PSNR and bitrate. In the future work, we would like to apply the proposed statistical learning approach to other H.264 components, like block matching algorithm, the computation of RDcost and the best mode selection between inter and intra modes. Another direction is to investigate more reliable features. In our experiments, we used median-motion sequences as our training samples. We would like to include a wide variety of videos of different motion patterns, such as fast, median and slow motions, into the training data to improve the SVM classification accuracy for different types of videos.

Acknowledgments. This work was supported in part by the National Science Council, Taiwan, R.O.C., under grant 97-2220-E-007-007.

References

1. Pan, W.-H., Chiang, C.-K., Lai, S.-H.: Fast Intermode Decision via Statistical Learning for H.264 Video Coding. In: Satoh, S., Nack, F., Etoh, M. (eds.) MMM 2008. LNCS, vol. 4903, pp. 329–337. Springer, Heidelberg (2008)
2. Chiang, C.-K., Lai, S.-H.: Fast Multi-Reference Motion Estimation Via Statistical Learning For H.264. In: IEEE International Conference on Multimedia & Expo. (ICME), New York (2009)
3. Hwang, C., Lai, S.-H.: Efficient Intra Mode Decision Via Statistical Learning. In: Ip, H.H.-S., Au, O.C., Leung, H., Sun, M.-T., Ma, W.-Y., Hu, S.-M. (eds.) PCM 2007. LNCS, vol. 4810, pp. 148–157. Springer, Heidelberg (2007)
4. Zhan, B., Hou, B., Sotudeh, R.: A Novel Fast Inter Mode Decision Algorithm Based On Statistic And Adaptive Adjustment For H.264/AVC. In: International Conference on Software, Telecommunications and Computer Networks, SoftCOM, pp. 1–5 (2007)

5. Huang, Y.-W., Hsieh, B.-Y., Wang, T.-C., Chen, S.-Y., Ma, S.-H., Shen, C.-F., Chen, L.-G.: Analysis And Reduction Of Reference Frames For Motion Estimation In MPEG-4 AVC/JVT/H.264. In: Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (2003)
6. Ma, W., Yang, S., Gao, L., Pei, C., Yan, S.: Fast Mode Selection Scheme For H.264/AVC Inter Prediction Based On Statistical Learning Method. In: IEEE International Conference on Multimedia and Expo., ICME (2009)
7. Wu, P., Xiao, C.-B.: An Adaptive Fast Multiple Reference Frames Selection Algorithm For H.264/AVC. In: Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (2008)
8. Sim, D.-G., Kim, Y.: Context-adaptive Mode Selection For Intra-block Coding In H.264/MPEG-4 Part 10. Real-Time Imaging 11, 1–6 (2005)
9. Joshi, U., Jillani, R., Bhattacharya, C., Kalva, H., Ramakrishnan, K.R.: Speedup Macroblock Mode Decision In H.264/SVC Encoding Using Cost-sensitive Learning. In: Digest of Technical Papers International Conference on Consumer Electronics, ICCE (2010)
10. Cortes, C., Vapnik, V.: Support-Vector Networks. Machine Learning, 273–297 (1995)