A Machine Learning Approach to Reducing Image Coding Artifacts

Ichiro Matsuda, Tomokazu Ishikawa, Yusuke Kameda, and Susumu Itoh

Department of Electrical Engineering, Faculty of Science and Technology,

Tokyo University of Science

2641 Yamazaki, Noda-shi, Chiba 278-8510, JAPAN

Email: matsuda@ee.noda.tus.ac.jp

Abstract—In this paper, a method for reducing coding artifacts introduced by lossy image compression is proposed. The method is similar to sample adaptive offset (SAO) which is adopted in the H.265/HEVC video coding standard as one of in-loop filtering tools. In the SAO, samples of the reconstructed image are classified into several categories based on some simple algorithms, and an optimum offset value is then added to the samples belonging to each category. Since the classification algorithms are switched on a block-by-block basis, not a negligible amount of side-information must be transmitted to the decoder in addition to the offset values. On the other hand, our method adopts a machine learning technique using a support vector machine (SVM) for the classification process. By applying the common SVM classifier to a whole image, the amount of the side-information can be considerably reduced. Simulation results indicate that the proposed method provides bitrate savings of up to 1.0 % for HD size images degraded through intra frame coding of the H.265/HEVC standard.

Index Terms—Post filtering, coding artifacts, machine learning, support vector machine (SVM), sample adaptive offset (SAO)

I. Introduction

Post filtering, or in-loop filtering for video coding, is a technique to recover image quality degraded through lossy image compression, and recent coding standards employ this kind of technique to achieve better coding efficiency. For example, a deblocking filter, which reduces blocking artifacts caused by block-based motion-compensated prediction and/or transform coding, is adopted in H.264/AVC video coding standard [1]. In development of the latest H.265/HEVC standard [2], three kinds of in-loop filtering tools: improved deblocking filter (DF) [3], adaptive loop filtering (ALF) [4] and sample adaptive offset (SAO) [5] were investigated, and finally DF and SAO have been accepted. While DF and ALF follow a typical filtering procedure where filter output is generated as weighted sum of multiple reconstructed values in local neighbors, the SAO utilizes the local information to merely classify a target sample into several categories. The filtering output is then obtained by adding an optimum offset value to samples corresponding to each category. It can be seen as a pattern analysis approach and classification algorithms play an important role in obtaining better coding gains. In practice, the SAO has several classification algorithms that are designed in advance and the best one in a rate-distortion sense is selected for each fixed size block, called a coding tree unit (CTU). It means that not a negligible amount of side-information on selection of the algorithms must be transmitted to the decoder in addition to the offset values.

Recently, a novel nonlinear post filtering technique has been proposed in [6], where a filtering algorithm itself is designed for each image using a evolutionary computing framework. It requires high computational power at the encoder side and the resulting algorithm must also be transmitted to the decoder as side-information. In [7], a support vector machine (SVM), which is known as a robust supervised machine learning algorithm [8], is incorporated with a median filter for restoring images corrupted by impulse-like additive noise. Since the SVM is reputed to have high generalization ability, it is expected to be used for a wide variety of images without any extra side-information for the post filtering purpose.

In this paper, we propose a new post filtering method using the SVM for coding artifacts reduction. The method is similar to an edge offset (EO) mode of the SAO. Specifically, the sample classification algorithm in the EO mode is replaced by the SVM classifier which was trained on several decoded images in advance. After performing the SVM-based classification, an optimum offset value calculated at the encoder side is received as side-information and added to the samples belonging to each category. Since we employ a three-class SVM in this paper, only three small offset values are needed as the side-information for the post filtering process over the whole image.

II. Proposed Method

A. Motivation

In the H.265/HEVC standard, SAO has two kinds of classification modes: EO and band offset (BO), which are used exclusively for each CTU. In the EO, the reconstructed value at a target sample is compared with two of adjacent samples along a one dimensional directional pattern, which is called an EO class, to classify the target sample into one of five categories. Since the directional pattern can be rotated by $45°$, there are four EO classes to be signalized for each CTU. Though the parameters including the EO classes as well as offset values for the respective categories can be shared with adjacent CTUs, the amount of side-information required for the SAO is unable to disregard in general [5].

Our aim is to replace the above classification algorithms with a SVM classifier which can separate the samples so that each of the classified categories has a certain bias in reconstruction errors. If such a SVM classifier is commonly used independent of image contents, the side-information on the classification algorithms can be considerably reduced.

B. SVM Classifier

In order to capture local image structure around the target sample $\boldsymbol{p}_0$, reconstructed values of 13 samples shown in Fig. 1 are extracted in the proposed method. We assume that a local mean of image signals is irrelevant to statistical property of the reconstruction errors and, therefore, differences of the reconstructed values between these samples and the target one are used as input features for the SVM classifier. Moreover, we use a sigmoid-like function $S(d)$ to map the features within an interval of $[-1, +1]$ because it is usually recommended to scale the feature vector elements to a certain range [9]. Consequently, a 12 dimensional feature vector is defined with respect to the target sample $\boldsymbol{p}_0$ as follows:

$$\boldsymbol{x}(\boldsymbol{p}_0) = [\, x_1, x_2, \ldots, x_{12}\,]^t, \tag{1}$$

$$x_k = S\big(\tilde{f}(\boldsymbol{p}_k) - \tilde{f}(\boldsymbol{p}_0)\big), \tag{2}$$

$$S(d) = \frac{2}{1 + \exp(-a\cdot d)} - 1, \tag{3}$$

where $\tilde{f}(\boldsymbol{p}_k)$ represents the reconstructed value at the sample $\boldsymbol{p}_k$ and a is a gain factor which controls slope of the function $S(d)$ at $d = 0$ as shown in Fig. 2.

In this paper, we use a three-class SVM which predicts a class label $y(\boldsymbol{p}_0) \in \{-1, 0, +1\}$ from the above feature vector. As the desired output of the SVM used in supervised learning,
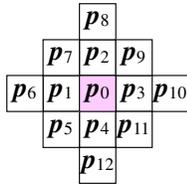

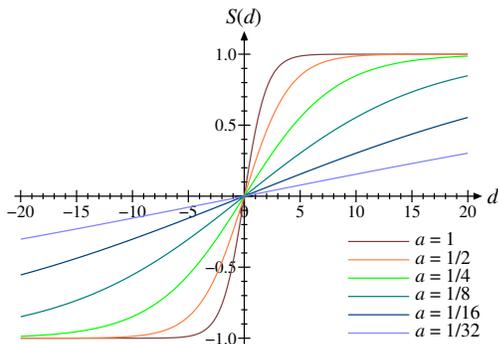Fig. 1. Target sample $\boldsymbol{p}_0$ and its neighbors.
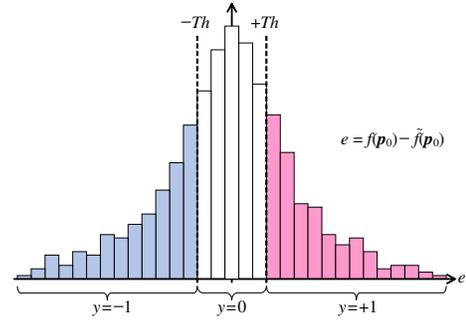

Fig. 2. Scaling function $S(d)$.

Fig. 3. Quantization of reconstructed errors to assign class labels.

an index of three-level quantization of the actual reconstruction error at each sample is assigned to the class label:

$$y(\boldsymbol{p}_0) = \begin{cases} -1 & \text{if} \quad f(\boldsymbol{p}_0) - \tilde{f}(\boldsymbol{p}_0) < -Th \\ +1 & \text{if} \quad f(\boldsymbol{p}_0) - \tilde{f}(\boldsymbol{p}_0) > +Th \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

where $f(\boldsymbol{p}_0)$ means the original image value and Th is a quantization threshold for the reconstruction errors. A value of Th is determined so that frequencies of the three quantization outputs can be equally distributed for the training image as depicted in Fig. 3

In this way, a training data set composed of pairs of the feature vector and the class label $\{\boldsymbol{x}(\boldsymbol{p}_0), y(\boldsymbol{p}_0)\}$ is collected from the given reconstructed images. By using this data set, the trained SVM classifier is expected to roughly discriminate the samples having significant biases in the reconstruction errors. Practically, we employ a nonlinear soft margin SVM with the radial basis function (RBF) kernel implemented in LIBSVM, an open source library for SVMs [10]. The following parameters may have an influence on the classification performance of the proposed method and their better settings will be investigated in the next section.

C: Regularization parameter for the misclassification errors.
γ: Kernel parameter used in the RBF.
a: Gain factor for the scaling function $S(d)$ defined in (3).

C. Post Filtering

The SVM classifier is applied to the target image containing coding artifacts, and each sample is categorized into one of three classes according to its output. Then the reconstruction errors obtained at the encoder side are averaged in each category and the optimum offset $\Delta f_y \in \mathbb{Z}$ is calculated by rounding off the averaged value into the nearest integer. Since three values of Δf_y $(y = -1, 0, +1)$ are usually close to zero, they can be transmitted to the decoder by a few bits. In this paper, we simply use the unary code [11] for the absolute values of Δf_y and add a sign bit if $\Delta f_y \neq 0$ as the side-information. It is worth noting that, our experimental results show that Δf_0 is always zero and signs of Δf_{-1} and Δf_{+1} are consistent with their class labels in almost all cases. These facts suggest that there is a room for further reducing the side-information, but its total amount is anyway much smaller

than the one needed for the SAO. Finally, the transmitted offset value is added to the reconstructed values of the samples belonging to each category. By carrying out this process for all the samples at the decoder side, the coding artifacts on the reconstructed image can be reduced.

III. EXPERIMENTAL RESULTS

A. Training and parameter tuning

In our experiments, the first frames of twelve CIF size (352×288 pels) monochrome sequences shown in Fig. 4 are encoded by the H.265/HEVC to obtain the training data set. Table I summarizes the coding condition used in this process. Appropriate values of the parameters C, γ and a mentioned in II-B are searched based on the leave-one-out cross-validation procedure, namely, when one image is being evaluated, the remaining eleven images are used for training of the SVM. To speed up this parameter tuning process, 10% of the samples are randomly drawn from the eleven images for constructing the training data set. Objective coding gains obtained by the proposed method are measured by Bjøntegaard delta bitrate (BD-rate) [12] calculated using four quantization parameters $QP = 22, 27, 32, 37$. This means three of the four QPs are different from the training condition shown in Table I.
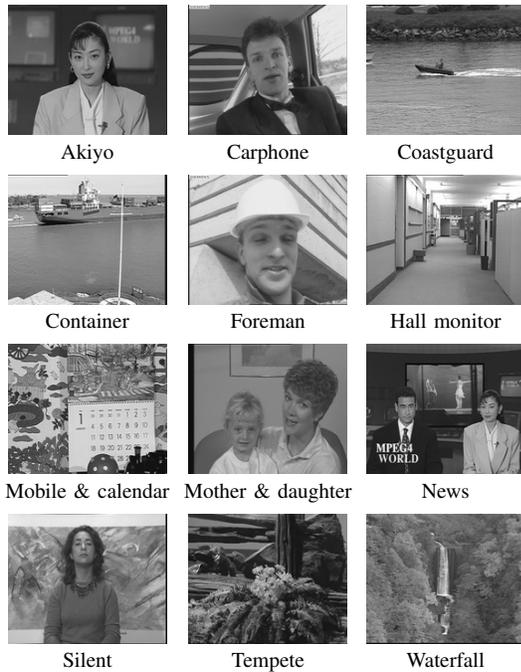

Fig. 4. Training images.

Akiyo   Carphone   Coastguard
Container   Foreman   Hall monitor
Mobile & calendar   Mother & daughter   News
Silent   Tempete   Waterfall

TABLE I
CODING CONDITION FOR TRAINING DATA SET.

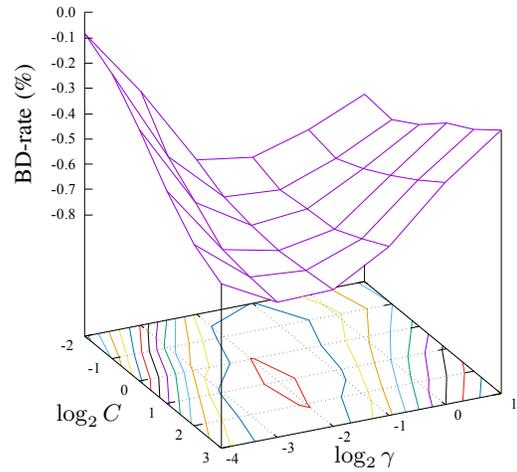| Codec	HM 16.7
Coding mode	Intra frame
Internal bit-depth	8 bit
Quantization parameter	Fixed (QP $=32$)
Deblocking filter (DF)	On
Sample adaptive offset (SAO)	Off

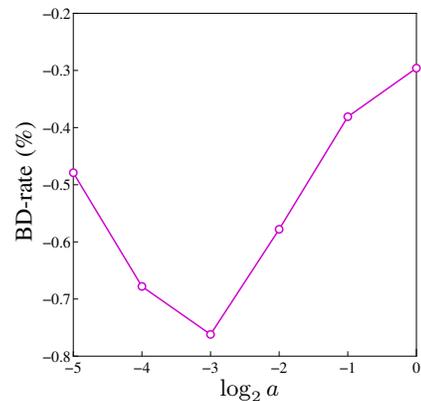Fig. 5. Relationship between BD-rates and the parameters C and γ.

Fig. 6. Relationship between BD-rates and the gain factor a.

Figure 5 shows a relationship between the averaged BD-rates and two parameters C and γ which are concerned with the soft margin SVM using the RBF kernel. According to [9], both of the parameters generally have a significant impact on the classification performance. In this two dimensional parameter search, the gain factor in (3) is fixed to $a = 0.125$. We can see that the combination of $C = 2.0$ and $\gamma = 0.25$ gives better coding gains on average. Under this condition, better choice of another parameter a is also searched. Figure 6 indicates that the setting of $a = 0.125$ is the best among the tested conditions. Based on these results, the SVM is finally trained using all of the twelve training images shown in Fig. 4 with the parameter settings of $C = 2.0$, $\gamma = 0.25$ and $a = 0.125$.

B. Performance evaluation

To eliminate bias from the training images, another set of images with HD resolution (1920×1080 pels, 8-bit grayscale) are used for performance evaluation. These images are the first frames of (a) TUM 1080p25 Data Set (No. 1 and 2) [13] and (b) ITE/ARIB Hi-Vision Test Sequence 2nd Edition (No. 201–210) [14], and their thumbnails are shown in Fig. 7.

TABLE II
COMPARISON OF BD-RATES.

Image	SAO	Proposed method	SAO + Proposed
Crowd run	-0.591%	-0.661%	-0.938%
Park joy	-0.663%	-0.484%	-0.871%
Ginkgo trees	-0.600%	-0.456%	-0.806%
Truck train	-0.519%	-0.257%	-0.554%
Cosmos flowers	-0.682%	-0.661%	-0.945%
Red leaves (pan up)	-0.550%	-0.393%	-0.693%
Sunlight through leaves	-0.825%	-1.015%	-1.314%
Red leaves (pan down)	-0.705%	-0.672%	-1.014%
Woman at harbor (circle dolly)	-0.802%	-0.481%	-0.873%
Fountain (follow)	-0.918%	-0.596%	-1.250%
Fountain (dolly)	-0.384%	-0.045%	-0.429%
Studio concert (confetti)	-0.771%	-0.801%	-1.230%
Average	-0.668%	-0.543%	-0.910%

Crowd run   Park joy
(a) TUM 1080p25 Data Set

Ginkgo trees   Truck train

Cosmos flowers   Red leaves (pan up)

Sunlight through leaves   Red leaves (pan down)

Woman at harbor (circle dolly)   Fountain (follow)

Fountain (dolly)   Studio concert (confetti)
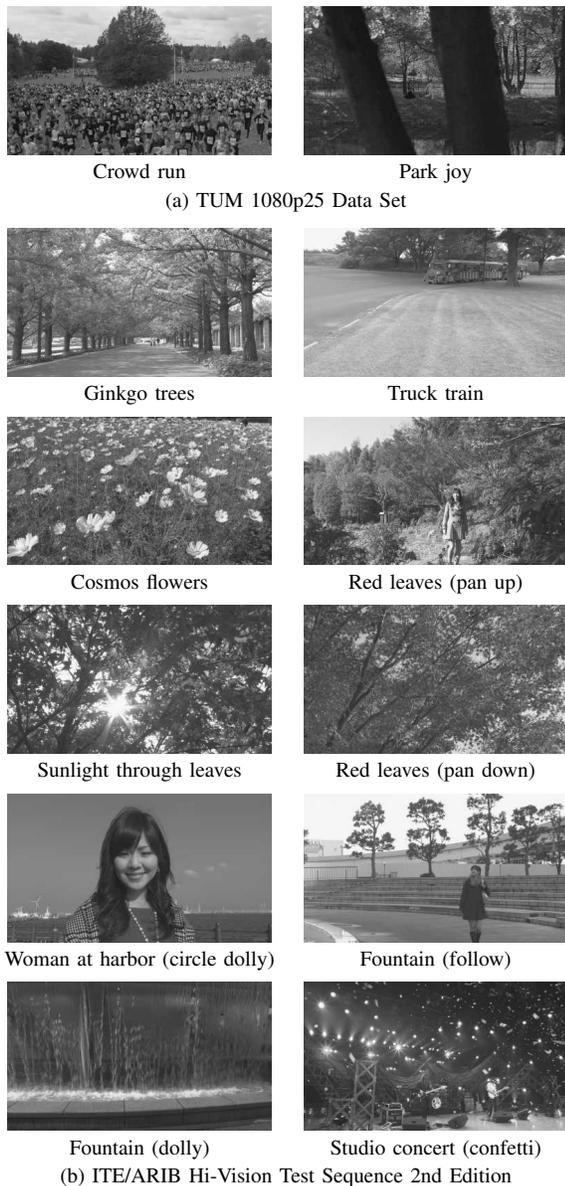(b) ITE/ARIB Hi-Vision Test Sequence 2nd Edition

Fig. 7. Test images used for the performance evaluation.

Intra frame coding of H.265/HEVC (without SAO) is used again as an anchor method in calculation of the BD-rates.

Table II compares the BD-rates obtained by the conventional SAO and the proposed method. This table also reports results of the proposed method incorporated with the SAO. In this case, the SAO is enabled in the training and post filtering processes, that is the SVM classifier is carried out after performing the SAO on the reconstructed images. We can see that, though the SAO provides better result on average. the maximum coding gain is achieved by the proposed method for "Sunlight through leaves". Furthermore, combination of both methods attains the best performance for all tested images and its averaged BD-rate reaches about -0.9%.

In Fig. 8, offset values calculated by the respective post filtering methods with $QP = 32$ are visualized by replacing its chroma channel (C_R component) with the offset values after $50\times$ amplification. It is observed that the offset values of the SAO tend to exhibit isolated patterns except for some blocks where BO is selected, while those of the proposed method form blob-like patterns. When both of the methods are incorporated, these patterns are overlaid and thus higher PSNR is achieved without losing their advantages.

IV. CONCLUSIONS

In this paper, we proposed a new post filtering method using a support vector machine (SVM) which is known as one of the most effective machine learning algorithms. In this method, the SVM is trained to classify the image samples into three categories so that the reconstructed errors introduced by lossy image coding have certain biases. After this classification, an optimum offset value which minimizes the reconstructed errors is simply added to the samples belonging to each category.

Effectiveness of the proposed method was evaluated using several images other than the training ones under the condition of intra frame coding of the H.265/HEVC. As a result, it was shown that the proposed method provides bitrate savings of up to 1.0% and can be used together with the state-of-the-art post filtering technique, namely sample adaptive offset (SAO). Since our method utilizes no prior knowledge of a specific

(a) Anchor method (33.75 dB, 0.542 bits/pel)

(b) SAO (33.81 dB, 0.543 bits/pel)

(c) Proposed method (33.82 dB, 0.542 bits/pel)

(d) SAO + Proposed (33.85 dB, 0.543 bits/pel)

Fig. 8. Enlarged view of "Crowd run" with visualization of the offset values.

coding algorithm, it could be used for any kinds of lossy image coding schemes as far as their coding artifacts have some structural relation to local waveform of the reconstructed images.

Currently, a major limitation of the proposed method is its complexity. In our experiments, the filtering process of a HD size image takes a few hours due to a large number of support vectors used in the classifier. A part of our future studies will be directed toward simplification of the classifier so that it can be used as an in-loop filtering tool for video coding.

REFERENCES

[1] ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC, "Advanced Video Coding for Generic Audiovisual Services", 2003.

[2] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, Y.-K. Wang and T. Wiegand, "High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & Last Call)," *JCT-VC, Doc. JCTVC-L1003*, Jan. 2013.

[3] A. Norkin, G. Bjøntegaard, A. Fuldseth, M. Narroschke, M. Ikeda, K. Andersson, M. Zhou and G. Van der Auwera, "HEVC Deblocking Filter", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp.1746–1754, Dec. 2012.

[4] C.-Y. Tsai, C.-Y. Chen, T. Yamakage, I. S. Chong, Y.-W. Huang, C.-M. Fu, T. Itoh, T. Watanabe, T. Chujoh, M. Karczewicz and S.-M. Lei, "Adaptive Loop Filtering for Video Coding", *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp.934–945, Dec. 2013.

[5] C.-M. Fu , E. Alshina, A. Alshin, Y.-W. Huang, C.-Y. Chen, C.-Y. Tsai, C.-W. Hsu, S.-M. Lei, J.-H. Park and W.-J. Han, "Sample Adaptive Offset in the HEVC Standard", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp.1755–1764, Dec. 2012.

[6] S. Takamura and A. Shimizu, "GPGPU-Assisted Nonlinear Denoising Filter Generation for Video Coding", *Proc. of 2014 Genetic and Evolutionary Computation Conference (GECCO' 14)*, pp.151–152, July 2014.

[7] T.-C. Lin and P.-T. Yu, "Adaptive Two-Pass Median Filter Based on Support Vector Machines for Image Restoration", *Neural Computation*, vol. 16, no. 2, pp.333–354, Feb. 2004.

[8] B. Schölkopf and A. J. Smola, "Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond", MIT Press, 2001.

[9] C.-W. Hsu, C.-C. Chang and C.-J. Lin, "A Practical Guide to Support Vector Classification", http://www.csie.ntu.edu.tw/˜cjlin/papers/guide/guide.pdf

[10] C.-C. Chang and C.-J. Lin, "LIBSVM: a Library for Support Vector Machines". *ACM Trans. on Intelligent Systems and Technology*, vol. 2, no. 3, pp.2:27:1–27:27, Apr. 2011.

[11] S. W. Golomb, "Run-Length Encodings", *IEEE Trans. on Information Theory*, vol. 12, no. 3, pp.399–401, July 1966.

[12] G. Bjøntegaard, "Calculation of Average PSNR Differences between RD-curves", *ITU-T SG16 Q.6 Document VCEG-M33*, Apr. 2001.

[13] C. Keimel, J. Habigt, T. Habigt, M. Rothbucher and K. Diepold, "Visual Quality of Current Coding Technologies at High Definition IPTV Bitrates," *Proc. of 2010 IEEE International Workshop on Multimedia Signal Processing (MMSP 2010)*, pp.390–393, Oct. 2010.

[14] "ITE/ARIB Hi-Vision Test Sequence 2nd Edition Reference Manual", The Institute of Image Information and Television Engineers, Aug. 2011. https://www.ite.or.jp/contents/chart/emanual.pdf