OPTIMAL INTRA CODING OF HEVC BY STRUCTURED SET PREDICTION MODE WITH DISCRIMINATIVE LEARNING

Wenrui Dai, Hongkai Xiong

Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ABSTRACT

This paper proposes a novel model on intra-coding for high efficiency video coding (HEVC), which can simultaneously make the set of prediction for block of pixels in an optimal rate-distortion sense. It not only utilizes the spatial statistical correlation for the optimal prediction based on 2-D contexts, but also formulates the data-driven structural interdependencies to make the prediction error coherent with the probability distribution which is favorable for subsequent transform and coding. The so-called structured set prediction model incorporates max-margin Markov network to regulate and reason the multiple prediction in the blocks. The model parameters are learned by discriminating the actual pixel value from the other possible estimates to the maximal margin. Distinguished from the existing methods concerning the minimal prediction error, the Markov network is adaptively derived to maintain the coherence of set of prediction. To be concrete, the proposed model seeks the concurrent optimization of the set of prediction by relating the loss function to the probability distribution of subsequent DCT coefficients. The prediction error is demonstrated to be asymptotically upper bounded by the training error under the decomposable loss function. For validation, we integrate the proposed model into HEVC intra coding and experimental results show obvious improvement of coding performance in terms of BD-rate.

Index Terms— Structured set prediction model, intra coding, discriminative learning

1. INTRODUCTION

Video coding is the key technique for the applications from mobile terminals to the servers, such as multimedia services, broadcasting, and vide communications and storage. H.264/AVC [1], the state-of-the-art video coding standard developed jointly by ITU-T and ISO/IEC, has achieved a vital efficiency. The video coding standard is based on the traditional hybrid coding scheme which incorporates many state-of-the-art techniques to achieve outstanding coding performance. The improvements are mainly obtained by exploring statistical redundancies through various employments of improved intra and inter predictions, variable block motion estimation and multiple reference frames.

The improvement of intra coding is a key step in the enhancement of the global efficiency of video coding, because intra frames represent a large proportion in the bitrate in many applications. The intra coding in H.264/AVC involves four subsequent steps including intra prediction, transform, quantization and entropy coding. In intra prediction, pixels from the encoded neighboring blocks are utilized to predict current block, such that the spatial redundancies among spatially adjacent blocks can be reduced. Apart from the fixed 8×8 block size in previous coding scheme, H.264/AVC allows 9 intra prediction modes for 4×4 and 8×8 block and 4 modes for 16×16 block respectively.The prediction residual is obtained by subtracting the prediction block from the original block. And subsequently, discrete cosine transform is applied for such blocks progressively. Afterwards, the transform coefficients are quantized and entropy coded.

In 2005, High performance Video Coding (HVC [2]) has been initialized for further investigation where various modes are advocated to suit regions of different properties. In the corresponding Key Technology Area (KTA) software, a bidirectional intra prediction (BIP) and separable directional transforms are absorbed [5]. It is further refined in HEVC [3], the successor of H.264/AVC, where the macroblock up to 64×64 is considered with the hierarchical partition tree and the angular intra predictors with up to 34 prediction modes. Later, the combined intra prediction [4] is proposed to better exploit the spatial redundancies with both open-loop prediction and the closed-loop prediction.

Recently, learning-based methods for intra frame prediction are prevailing. [6] proposes the structured priority Belief Propagation (BP) based inpainting (IP) to exploit the intrinsic nonlocal and geometric regularity in video samples and take it as additional mode for intra prediction in H.264/AVC. The geometric regularity in block-based prediction is also considered in [7], which models the blocks of pixels through the graphical model associated to the directed acyclic graph, and consequently, estimates the mass function by considering the spatial dependency. Such learning based methods interrelate the set of prediction with the generated graphical model, but the generative methods like Markov random field does not simultaneously obtain the optimal solution for the varying video data.

In this paper, we propose the structured set prediction

model for the intra coding in H.264/AVC by both catering the context-based prediction to the underlying probabilistic distribution of transform and considering the structural interdependencies in a local region. Contrary to previous hybrid prediction schemes, discriminative learning is initiated to exploit the inherent statistical correlation to directly conditioned on the 2-D contexts and their corresponding prediction. The obtained prediction residual are in conformity with the underlying probabilistic distribution for subsequent transform. Furthermore, structural interdependencies are utilized to make the set of prediction in a correlated region coherent with the underlying probability distribution for subsequent transform-based coding. Thus, the derived transform coefficients tend to concentrated to low frequency domain.

To be concrete, the max-margin Markov network is proposed to simultaneously predict the region of correlated pixels in the sense of the optimal transform-based coding. The max-margin estimation is proposed for the individual pixel-wise prediction, where model parameters are trained directly conditioned on the obtained data, and subsequently, utilized to make the optimal context-based prediction. Such parameters are adopted to combine the class of feature functions to characterize the varying local statistics. To achieve the goal of reducing a coding rate, the loss function is designed to meet with the probability distribution of the subsequent DCT coefficients. Such that the coefficients derived from the loss-augmented inference are optimized to transmit to coding engine.

The rest of the paper is organized as follows. In Section 2, we generally describe the intra coding framework and the proposed structured set prediction model. In Section 3, we formulate the proposed model by deriving loss function and upper bound for prediction errors, and find the optimal solution for block-based prediction. Extensive experimental results are validated in Section 4 on both objective and visual quality. Finally, we draw the conclusion in Section 5.

2. THE PROPOSED FRAMEWORK

2.1. The Proposed CODEC

The generic video coding framework with the proposed structured set prediction model is depicted in Fig. 1. The proposed framework is based on HEVC Test Model under Consideration (TMuC). In addition to the existing intra- and inter-modes, each coding unit is designed to choose the optimal mode in the sense of the minimal rate-distortion cost. As marked in Fig. 1, the proposed structured prediction model is blended with the traditional angular intra prediction to serve as an alternative mode. Therefore, the value MODE_STRUCT is added to syntax element PRED_MODE.

$$\text{PRED\_MODE} \in \{\text{MODE\_SKIP}, \text{MODE\_INTER}, \text{MODE\_INTRA}, \textbf{MODE\_STRUCT}\}$$

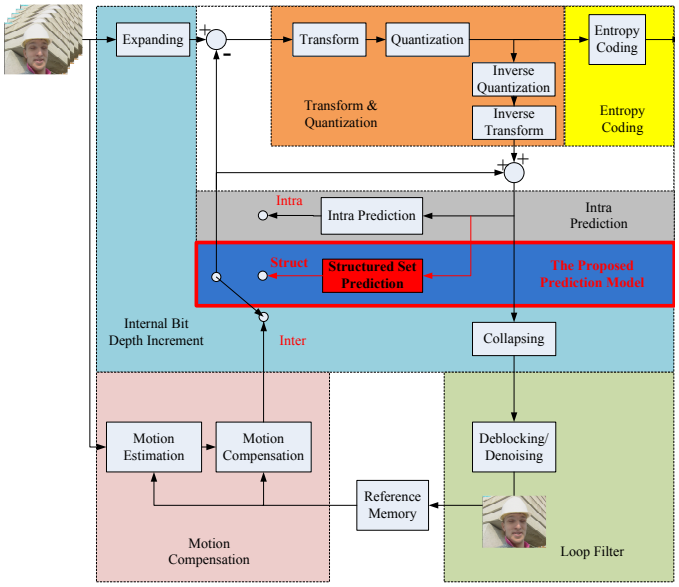


Fig. 1. The proposed codec with structured set prediction model based on TMuC. The structured set prediction model is blended with the original angular intra prediction to serve as an alternative prediction mode. The proposed model is selected according to the rate-distortion cost.

The STRUCT_MODE is initiated in intra(I) frame. The coding unit (CU) is iteratively predicted from the maximal possible size to the minimal one for the decision of prediction unit (PU) in intra prediction. After calculating the Lagrangian rate-distortion cost for all the possible intra prediction modes, the prediction units achieving the overall least cost is selected for the intra prediction of current coding unit. For each prediction unit, its PRED_MODE is chosen between MODE_INTRA and MODE_STRUCT, and the syntax element INTRA_PRED_MODE and PU_SIZE is accordingly derived in intra coding.

$$\text{INTRA\_PRED\_MODE} \in \{0, \cdots, 33\}$$
$$\text{PU\_SIZE} \in \{\text{PU\_4} \times 4, \text{PU\_8} \times 8, \text{PU\_16} \times 16, \text{PU\_32} \times 32, \text{PU\_64} \times 64\}$$

If P_n and P_{n-r} are the current and reconstructed prediction unit (PU) in the coding unit, the Lagrangian cost J_{P_n} of the P_n with parameter set $\text{PARAM} = \{\text{PRED\_MODE}, \text{PU\_SIZE}, \text{INTRA\_PRED\_MODE}\}$ is

$$J_{P_n}(P_n, \text{PARAM}|P_{n-r}, Qp, \lambda) = D(P_n, \text{PARAM}|P_{n-r}, Qp) + \lambda \cdot R(P_n, \text{PARAM}|P_{n-r}, Qp)$$

where Qp is the quantization parameter and λ is the Lagrange parameter associated with Qp. The Lagrange parameter λ is empirically set: $\lambda = 0.85 \cdot 2^{(Qp-12)/3}$. Similar to intra and inter modes, the predictor of struct mode is subtracted from the current prediction unit to generate the residue which

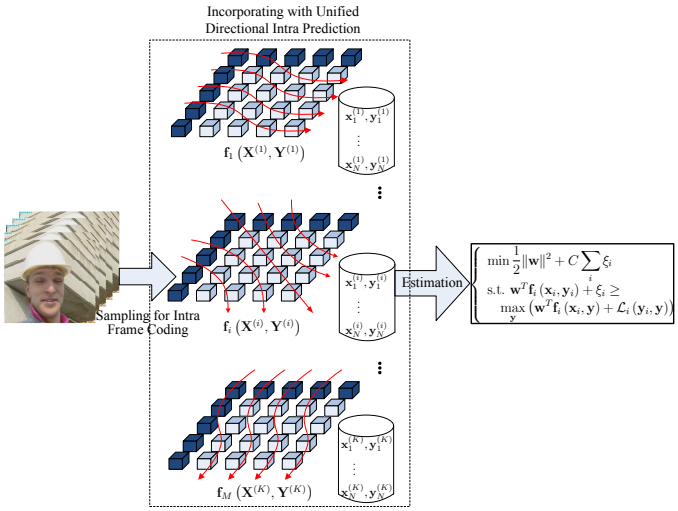


Fig. 2. Conceptual diagram for the structured set prediction model. The training data $\mathcal{S} = \{\mathbf{x}_i, \mathbf{y}_i\}$ for intra coding are dealt with the class of feature functions $\{\mathbf{f}_i\}$. The max-margin Markov network is trained over the training data by combining the class of feature functions $\{\mathbf{f}_i\}$ with the normal vector $\{\mathbf{w}_i\}$ for the max-margin estimation.

is subsequently transformed, quantized and encoded to obtain the compressed bitstream. No assistant side information needs to be included in the bit-stream, so that the proposed mode's rate R only contains the header information(including STRUCT_MODE flag) and the corresponding DCT residual.

2.2. The Structured Set Prediction Model

In this section, we describe the proposed structured set prediction model. Besides individual sequential prediction with adaptive rules for the varying spatial dependencies, the proposed model takes the correlation among the pixels for predicting into consideration, such that the prediction errors will adapt to the varying contexts. Consequently, the structured set prediction model aims both to make the context-based adaptive inference and to derive the constraints for sets of pixels for predicting. Fig. 2 illustrates the conceptual diagram for the structured set prediction model. In the proposed model, the concurrent training and prediction are performed for the block of pixels with the fixed size, such that enforced constraints on the interdependencies within the pixels can be learned and inferred to help improve the predictive performance. The proposed model is based on the class of feature functions $\mathcal{F} = \{\mathbf{f}_i(\mathbf{x}, \mathbf{y})\}$. The feature functions establish the conditional probabilistic model for prediction based on the various contexts derived from the supposed structural interdependencies.

$$\mathbf{f}_i(\mathbf{x}, \mathbf{y}) = P(\mathbf{y}|\mathbf{x}),$$

Consequently, the local spatial statistics is characterized by the linear combination of the class of feature functions.

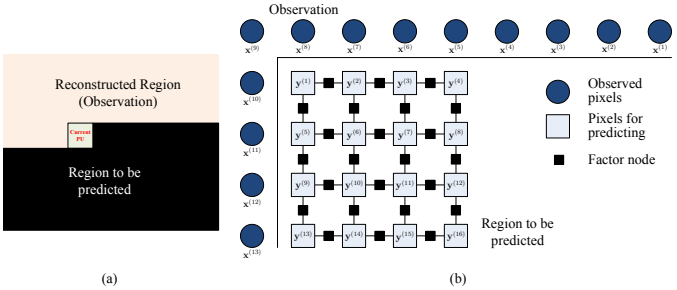


Fig. 3. The generated graphical model for the structured set prediction model, where $\{\mathbf{y}^{(i)}\}$ is the set of pixels to be predicted, $\{\mathbf{x}^{(i)}\}$ is the set of observed pixels serving as contexts

Fig. 3(a) shows the construction of the proposed structured set prediction model. Similar to the common intra prediction methods, the prediction is based on the adjacent reconstructed regions of the frame and the set of pixels are obtained simultaneously. Denoting $\mathbf{y}$ the set of pixels to be predicted and $\mathbf{x}$ the reconstructed pixels as contexts, its prediction is derived in the concurrent form.

$$\hat{\mathbf{y}} = \arg\max_{\mathbf{y}} \mathbf{w}^T \mathbf{f}(\mathbf{x}, \mathbf{y}) \tag{1}$$

where $\mathbf{f}$ is the collection of feature functions indicating the probability distribution conditioned on the various spatial structural interdependencies and $\mathbf{w}$ is the trained weighting vector of the linear model that combines the class of feature functions. The training process of the weighting vector $\mathbf{w}$ is modeled as an optimized problem considering both the context-based spatial correlation and the interdependencies among the pixels for predicting.

2.3. Intra Prediction as an Optimized Problem

It is reasonable to conclude that the predictive performance is essentially based on the training of weighting vector $\mathbf{w}$. Denote $\mathcal{S} = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{N}$ the collected set of training data, where $\mathbf{y}_i$ is the ith labeled block of pixels for predicting and $\mathbf{x}_i$ is the ith observed contexts for $\mathbf{y}_i$. Consequently, the training of weighting vector $\mathbf{w}$ is proposed to be an optimized problem over the training set $\mathcal{S}$. The min-max formulation of the max-margin Markov network is proposed for the training process by simultaneously considering the constraints representing the structural interdependencies among the pixels for predicting.

Since the pixels for predicting are naturally correlated with the mutual structural interdependencies in local regions, the min-max formulation is constructed according to the graphical model described in Fig. 3(b). If we suppose that the block of M pixels for predicting $\mathbf{y} = \{\mathbf{y}^{(j)}\}_{j=1}^{M}$ is also correlated for local coherence, the 2-D Markov network is constructed accordingly where each edge clique contains the two neighboring pixels connected by the edge. As a result, we make the max-margin estimation for each pixel and obtain the joint optimized prediction over all pixels by utilizing

the correlations in their neighborhood.

$$\begin{cases} \min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i \\ \text{s.t.} \quad \mathbf{w}^T\mathbf{f}_i(\mathbf{y}_i) + \xi_i \geq \max_{\mathbf{y}} \left(\mathbf{w}^T\mathbf{f}_i(\mathbf{y}) + \mathcal{L}(\mathbf{y}_i, \mathbf{y})\right) \quad \forall i \end{cases} \tag{2}$$

In (2), the weighting vector $\mathbf{w}$ is the normal vector perpendicular to the hyperplane spanned by the class of feature functions $\{\mathbf{f}_i\}$, and $\{\xi_i\}$ is the slack vector which allows for the violations of the constraints at a cost proportional to $\{\xi_i\}$. In view of the fact that practical coding is based on the probabilistic estimation of errors with the alleged distribution, the loss function $\mathcal{L}(\mathbf{y}_i, \mathbf{y})$ for measurement is defined to relate with the actual code length under such distribution. In consequence, the optimized problem is conducted under the class of feature functions $\mathcal{F} = \{\mathbf{f}_i(\mathbf{x}, \mathbf{y})\}$ and the loss function $\mathcal{L}(\mathbf{y}_i, \mathbf{y})$ measuring the actual code length.

3. FORMULATION OF STRUCTURED SET PREDICTION MODEL

3.1. Loss Function

Since there exists strong connection between the loss-scaled margin and the expected risk of the learned model, we are to make a study of the loss function for the loss-augmented inference. Given the M-ary estimated output $\hat{\mathbf{y}}$, its approximation error is supposed to be measured by the loss function $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$ that is defined jointly on the cliques of the generated graphical model. The construction of the loss function considers both the Laplacian errors derived for each node potential and the state transition of the neighboring nodes for each edge potential.

$$\begin{aligned} \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y}) = &\sum_i \ell_i\left(\hat{\mathbf{y}}^{(i)} - \mathbf{y}^{(i)}\right) \\ &+ \sum_i \sum_{j \in \text{ne}(i)} \mathbb{I}\left(\hat{\mathbf{y}}^{(i)}, \mathbf{y}^{(i)}\right) \mathbb{I}\left(\hat{\mathbf{y}}^{(j)}, \mathbf{y}^{(j)}\right) \end{aligned} \tag{3}$$

where $\ell_i(\cdot)$ is the Laplacian loss function for the ith node component based on the disparity between the ith label $\mathbf{y}^{(i)}$ and the its estimation $\hat{\mathbf{y}}^{(i)}$. In (3), we denote $\epsilon_i = \hat{\mathbf{y}}^{(i)} - \mathbf{y}^{(i)}$ the ith prediction error in set prediction and σ^2 the variance derived by the all M errors $\{\epsilon_i\}_{i=1}^M$ in the predictive region.

The reason for adoption of the Laplacian loss function is that the Laplace distribution is demonstrated to fit the 2D-DCT coefficients [8]. Consequently, the prediction errors measured with the Laplacian loss function meet with the practical DCT transform-based coding with least empirical entropy. Such that, the loss function is designed not only to indicate the disparities of the actual pixel values from the predicted ones but to meet with the 2D-DCT transform for a concentrated DC energy.

$$\ell_i(\epsilon_i) = \begin{cases} \log_2\left(1 - e^{-\frac{1}{\sqrt{2}\sigma}}\right) & \epsilon_i = 0 \\ \log_2\left(\frac{1}{2}\left(e^{-\frac{|\epsilon_i|-0.5}{\sigma/\sqrt{2}}} - e^{-\frac{|\epsilon_i|+0.5}{\sigma/\sqrt{2}}}\right)\right) & 0 < |\epsilon_i| < 255 \\ \log_2\left(\frac{1}{2}e^{-\frac{|\epsilon_i|-0.5}{\sigma/\sqrt{2}}}\right) & |\epsilon_i| = 255 \end{cases}$$

where e is the base of the natural logarithms. Consequently, the solutions to the loss-augmented optimization problem will minimize the practical code length under $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$.

3.2. Upper Bound for the Prediction Errors

In this section, we show that the upper bound for prediction error is asymptotically equivalent to the training error. Such upper bound allows us to relate the error on the training data to the prediction error. Consequently, the prediction performance is assured to converge as long as the weighting vector $\mathbf{w}$ has been well-tuned to fit the training data. As mentioned above, the proposed model aims to minimize the cumulative code length of a correlated region in terms of Laplacian measurement. Extending the average error $\mathbf{L}(\mathbf{w} \cdot \mathbf{f}, \mathbf{y})$ for the blocks of M pixels to measure with the γ-margin hypersphere, we define a γ-margin per-label loss.

$$\mathbf{L}^\gamma(\mathbf{w} \cdot \mathbf{f}, \mathbf{y}) = \sup_{\mathbf{y}': \|\mathbf{w}\cdot\mathbf{f}(\mathbf{y}) - \mathbf{w}\cdot\mathbf{f}(\mathbf{y}')\| \leq \gamma\mathcal{L}(\mathbf{y}, \mathbf{y}')} \frac{1}{M}\mathcal{L}(\mathbf{y}, \mathbf{y}').$$

The γ-margin per-label loss $\mathbf{L}^\gamma(\mathbf{w} \cdot \mathbf{f}, \mathbf{y})$ picks from all proper $\mathbf{y}'$ (satisfies $\mathbf{w}^T\mathbf{f}(\mathbf{x}, \mathbf{y}) \leq \mathbf{w}^T\mathbf{f}(\mathbf{x}, \mathbf{y}')$) that maximizes the log-Gaussian measure from $\mathbf{y}$ in a $\gamma\mathcal{L}(\mathbf{y}, \mathbf{y}')$ wider hypersphere. It is actually closed to the loss in the proposed min-max formulation. We prove that the prediction and training is asymptotically consistent, which means that the upper bound for prediction error will converge to the training error with sufficient sampling.

Proposition 1. *For the trained normal vector* $\mathbf{w}$ *and arbitrary constant* $\eta > 0$*, the prediction error is asymptotically equivalent to the one obtained over the training data with probability at least* $1 - e^{-\eta}$.

Actually the prediction error is upper bounded by two additional terms. The first term bounds the training error based on $\mathbf{w}$. The low training error $\mathbb{E}_{\mathcal{S}}\mathbf{L}^\gamma(\mathbf{w} \cdot \mathbf{f}, \mathbf{y})$ can be achieved with the well-tuned weighting vector $\mathbf{w}$. Such that the prediction model can be upper bounded by the low error $\mathbb{E}_{\mathcal{S}}\mathbf{L}^\gamma(\mathbf{w} \cdot \mathbf{f}, \mathbf{y})$ and high margin γ. The second term is the excess loss corresponding to the complexity of the predictor, which vanishes with the growth of sample size N. As a result, the expected average predictive error is asymptotically equivalent to the γ-margin per-label error.

Proposition 1 ensures the predictive performance by relating the theoretical upper bound for prediction to the tunable training error. Since the loss derived by the Laplacian

loss function meets with the empirical distribution of DCT coefficients [8], the average loss can be related to the practical code length of prediction residue. In intra frame video coding, it is noted the practical coding length led by the proposed model asymptotically approaches the well-tuned discriminative results over the sufficient sampling data from the underlying distribution. Consequently, structured set prediction model minimizes the coding length of prediction residue to the well-tuned loss over the training data. Considering the characterization of the learning-based methods, the prediction tends to be efficient in the regions with regular features. As a result, the coding performance in such regions tends to be better as the DCT coefficients shall be concentrated under the log-Laplacian loss function.

Algorithm 1 Message passing with expectation propagation

Initialize $\tilde{\psi}_C(\mathbf{x})$ for all cliques and make $q(\mathbf{x})$ their product
repeat
  for all C **do**
    $q^{\setminus\psi_C}(\mathbf{x}) = q(\mathbf{x}) / \tilde{\psi}_C(\mathbf{x})$
    $q'(\mathbf{x}) = \mathcal{P}\left[\psi_C(\mathbf{x}) q^{\setminus\psi_C}(\mathbf{x})\right]$
    $\tilde{\psi}_C(\mathbf{x})^{\text{new}} = \tilde{\psi}_C(\mathbf{x}) \left(\frac{q'(\mathbf{x})}{q(\mathbf{x})}\right)$
    $q(\mathbf{x})^{\text{new}} = q'(\mathbf{x})$
  end for
until Convergence

3.3. Solving Structured Set Prediction Model

Since the standard quadratic programming (QP) for solving Eq. (2) is often prohibitive in the structured set prediction model even for small training sets. We obtain its dual to use the coordinate dual ascent method analogous to the sequential minimal optimization (SMO). SMO breaks the optimization problem into a series of small QP problems and takes an ascent step that modifies the least number of variables.

$$\begin{cases} \max\left[v_i(\mathbf{y}') - v_i(\mathbf{y}'')\right]\delta - \dfrac{1}{2}C\|\mathbf{f}_i(\mathbf{y}') - \mathbf{f}_i(\mathbf{y}'')\|^2\delta^2 \\ \text{s.t.}\alpha_i(\mathbf{y}) + \delta \geq 0, \alpha_i(\mathbf{y}'') - \delta \geq 0 \end{cases} \tag{4}$$

where $v_i(\mathbf{y}) = \mathbf{w} \cdot \mathbf{f}_i(\mathbf{y}) + \mathcal{L}(\mathbf{y}_i, \mathbf{y})$. The minimization process chooses the SMO pairs with respect to the KKT conditions. The KKT conditions are the sufficient and necessary criteria for optimality of the dual solution, which commits the certain locality for each example. The detailed algorithms for SMO can be referred to [9].

To maintain the spatial structures, we build the grid-like Markov random field (shown in Fig. 3(b))for the α_i and v_i in each block, and accordingly calculate their marginal to decide the SMO pairs. Since the generated Markov random field is not a chordal graph, it is firstly triangulated into a corresponding junction tree for cliques which can be obtained. As the triangulation is not necessarily unique, we choose the chain-like junction tree for simplicity. Subsequently, the inference algorithm to find the states achieving the maximal probability is applied. It involves two points: the selection of divergence measure and the message-passing scheme. In this section, we propose the expectation propagation (EP) based methods for seeking the most probable states. As an extension to belief propagation, expectation propagation can solve the problems without explicit posterior distribution over a single variable for it sends only expectations of features in message-passing.

Expectation propagation estimates cumulative distribution of the grid-like graphical model by approximating the factors one by one. For clique C, the self-excluded estimated marginal $q^{\setminus\psi_C}$ is calculated.

$$q^{\setminus\psi_C}(\mathbf{x}) = q(\mathbf{x}) / \tilde{\psi}_C(\mathbf{x})$$

The estimated cumulative distribution $q^{\setminus\psi_C}/\psi_C$ is approximated

$$q'(\mathbf{x}) = \mathcal{P}\left[\psi_C(\mathbf{x}) q^{\setminus\psi_C}(\mathbf{x})\right]$$

$\mathcal{P}[\cdot]$ chooses functions from the exponential family in the sense of KL-divergence. The potential ψ_C and the product of cliques q is then updated and passed to next clique. Such approximation and update process is propagated over the junction tree. After the local propagation in each junction is finished, the whole state space can be obtained by combining their results. Such that expectation propagation is able to seek the min-max formulation over the generated junction tree. For the clique C, its estimation and update with the message passed in and out are described in Algorithm 1.

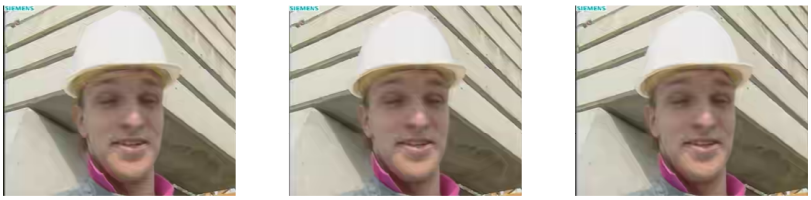
4. EXPERIMENTAL RESULTS

The proposed model based intra prediction scheme is implemented based the HEVC test model TMuC 0.9 [10]. The proposed model is integrated into the intra coding with the hierarchical tree-structured PU sizes ranging from 64×64 to 4×4. Six test sequences with various resolutions, including two CIF (352×288), one WQVGA (416×240), one standard definition (832×480), one 720p definition (720×576) and one high-definition (1920×1080), are evaluated. All the test sequences are encoded with the same quantization parameters for fair comparison, as the techniques adopted in video coding would be interfered by the rate control algorithm. The test conditions used for evaluation is can be referred to [12].

Table 1 show the BD-rates [11] of the test video sequences, where the combined intra prediction is selected as the benchmark. The maximum coding unit size is fixed to 64 and the maximum partition size is 4, such that the size of prediction unit can be ranged from 64×64 to 4×4. The rate-distortion point is obtained respectively from coding result with various QP (24, 28, 32, 36) without rate control. Table 1 shows that the proposed model obtains approximately 1-2% average bit rate reduction compared with the angular intra prediction in TMuC 0.9 reference software. Fig. 4 shows the reconstructed videos based on the three methods and the

Table 1. Results for BD-rate comparison with the combined intra prediction

Sequence	Size	The proposed model			Combined intra prediction		
		Y BD-rate	U BD-rate	V BD-rate	Y BD-rate	U BD-rate	V BD-rate
Akiyo	352×288	2.08	0.11	0.20	1.09	2.11	1.77
Foreman	352×288	2.68	0.75	0.02	0.87	2.17	2.22
BQSquare	416×240	1.22	0.43	0.25	0.62	0.76	0.05
Basketball	720×576	1.44	0.38	0.48	0.75	0.77	0.78
BQMall	832×480	1.08	0.05	0.13	1.05	1.07	1.20
BQTerrace	1920×1080	1.46	0.88	1.75	0.78	0.77	0.08



(a) (b) (c)

Fig. 4. Comparative results for Foreman with the (a) proposed model, (b) CIP, (c) TMuC at QP=36

proposed model is found to obtain comparative visual quality when compared with CIP.

5. CONCLUSION

The learning-based structured set prediction model for intra coding is proposed in this paper. The proposed model makes set prediction for a correlated region of pixels simultaneously by considering both the inherent statistical correlation for context-based prediction and structural interdependencies for coherence of set prediction. Such that the obtained prediction error not only achieves the optimal solution for the individual context-based prediction, but also is optimized with the global structural interdependencies. The training and prediction are formulated with the max-margin Markov network, where the collective discriminative optimization is achieved under the well-defined loss function. With the growth of sample size, the loss-augmented inference is consistent with training in the sense of the concentration of the DCT coefficients. The proposed min-max formulation can be solved by combining the optimized results of all the individual cliques with expectation propagation for the lower dimension of the state spaces. In practice, the proposed model is integrated with the latest HEVC encoder and measured in terms of BD-rate.

6. REFERENCES

[1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560-576, Jul. 2003.

[2] ITU-T Q6/16 and ISO/IEC JTC1/SC29/WG11, "Vision, Applications and Requirements for High-Performance Video Coding (HVC)," in *ISO/IEC JTC1/SC29/WG11 MPEG, N11096*, (Kyoto, Japan), Jan 2010.

[3] ISO/IEC, "Joint call for proposals on video compression technology," in *ISO/IEC JTC1/SC29/WG11 MPEG N11113*, (Kyoto, Japan), Jan 2010.

[4] A. Gabriellini, D. Flynn, M. Mrak, and T. Davies, "Combined intra-prediction for high-efficiency video coding," *IEEE J. Sel. Topcis Signal Process.*, vol. 5, no. 7, pp. 1282-1289, Nov. 2011.

[5] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning," in *Proc. Int. Conf. Image Process.*, San Diego, CA, Oct. 2008, pp. 2116-2119.

[6] H. Xiong, Y. Xu, Y.-F. Zheng, and C. Chen, "Priority belief propagation based inpainting prediction with tensor voting projected structure in video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 8, pp. 1115-1129, Aug. 2011.

[7] S. Milani, "Fast H.264/AVC FRExt intra coding using belief propagation," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 121-131, Jan. 2011.

[8] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661-1666, Oct. 2000.

[9] B. Taskar, "Learning structured prediction models: A large margin approach," Ph. D. dissertation, Stanford Univ., CA, Dec. 2004. [Online]. Available: http://robotics.stanford.edu/btaskar/pubs/thesis.pdf/

[10] "HEVC Test Model under Consideration," [Online]. Available: http://hevc.kw.bbc.co.uk/

[11] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-Curves,", ITU-T SG16/Q6, 13th VCEG Meeting, Doc. VCEG-M33, Austin, Texas, USA, Apr. 2001.

[12] T. Wiegand, B. Bross, W.-J. Han, J.-R. Ohm, and G. J. Sullivan, "WD3: Working draft 3 of high-efficiency video coding," in JCTVC-E603, (Geneva, Switzerland), Mar. 2011.

[13] F. Bossen, "Common test conditions and software reference configurations," in JCTVC-D600, (Daegu, Korea), Jan. 2011.