INTRA PREDICTION USING TEMPLATE MATCHING WITH ADAPTIVE ILLUMINATION COMPENSATION

Yunfei Zheng†‡∗, Peng Yin†, Óscar Divorra Escoda†, Xin Li‡ and Cristina Gomila†

†Thomson Corporate Research, Princeton, New Jersey 08540, USA
‡West Virginia University, Department of CS&EE, Morgantown, West Virginia 26506, USA

ABSTRACT

Modern video coding standards such as H.264/AVC use intra-prediction for efficient coding of Intra pictures. These usually exploit local directional signal correlations. More recently, intra-prediction modes using non-local signal information have been introduced. A very popular approach is the so-called Template Matching Prediction (TMP), which uses template based texture synthesis for signal prediction. This, combined with regular directional prediction, significantly improves intra coding efficiency compared to H.264/AVC. However, current TMP techniques have trouble synthesizing picture data with non-uniform illumination characteristics. They assume that similar picture regions resemble at the same time in structure and illumination, which is often not the case. In order to solve this, we propose a template matching technique with locally adaptive illumination compensation. The proposed technique is based on a linear compensation model with a scaling and an offset parameters to compensate for contrast and brightness disparities respectively. The model parameters are calculated by solving an auto-regressive Least Square problem during the template search for TMP. This permits to synthesize signal structures while capturing the local characteristics of illumination without needing extra side information. The total improvement in intra coding efficiency with respect to H.264/AVC can be of up to 19%.

Index Terms— Video Coding, Intra Prediction, H.264/AVC, Template Matching, Illumination Compensation

1. INTRODUCTION

H.264/AVC Intra prediction is conducted by extrapolating the surrounding previously-decoded pixels of a predicted block. Eight directional prediction modes together with a DC mode are used to predict directional structures (e.g., edges at various angles) as well as smooth areas [1]. Despite their good performance on smooth and simple structured areas, H.264/AVC Intra prediction modes are only able to exploit local picture redundancy, leaving still room for improvement on more complicate signal structures such as textures.

Recently, some non-local approaches have been introduced for intra coding of pictures which show promising performance. Examples of these are Intra Displacement Compensation (IDC) [2] and Template Matching Prediction (TMP) [3]. These approaches try to exploit non-locally existing self-similarities within a picture. Motivated by inter motion compensation, IDC reuses block patches in the reconstructed decoded area of a picture in order to predict the current block. An intra displacement vector per block or partition is thus sent to the decoder as overhead (refer to Fig.1). As shown in Fig. 2, TMP is also generated by reusing available reconstructed data at the decoder. Unlike IDC, TMP uses backward-adaptive texture synthesis [4] techniques, requiring no overhead to be sent. Indeed, TMP measures the similarity between the surrounding neighboring pixels (available at both encoder and decoder) and candidate template blocks for prediction rather than the original block data as used in IDC. Since no additional overhead is required by TMP, the target block can be partitioned into smaller blocks for prediction. This allows for more accurate modeling of high frequency components and complicate structures. Furthermore, multiple prediction candidates can be fused to get a better prediction [5]. This is often not possible in IDC since the cost of coding multiple displacement vectors is not worth.

TMP approaches typically use the Sum of Absolute Differences (SAD) or Sum Squared Errors (SSE) between two templates in order to retrieve those with higher similarity. Although such searching scheme works well in most cases, it is not efficient enough when there exist some mismatch between templates. This may happen in the presence of illumination disparities or geometric variations, leading to sub-optimal prediction synthesis and a larger residue. Regular TMP approaches assume that similar picture regions resemble at the same time in structure and illumination, which is often not the case. Indeed, features with high local correlation like contrast and brightness are not always consistent with the non-local structural similarities of two different picture regions. In regular non-uniform illumination conditions, non-local information can just serve as a partial source of prediction data. In such case, local and non-local information should be combined to work out a more efficient prediction.

In this paper, we propose a TMP approach with adaptive illumina-
nation compensation. Borrowing ideas from weighted inter prediction in order to compensate temporal illumination variations and/or fade in/out effects [6, 7], we propose a linear compensation model with a scaling and an offset parameters to compensate contrast and brightness disparities in Intra TMP. In order to capture local illumination characteristics, the linear model parameters are adaptively calculated per template by solving a Least Square problem. This calculation is based on the template data which is available at both encoder and decoder, requiring no additional overhead to be sent. Such illumination compensation can also be applied to other non-local prediction approaches such as IDC.

This paper is organized as follows: TMP is first introduced in Sec. 2. Then, the proposed adaptive illumination compensated TMP scheme is described in Sec. 3. Sec. 4 illustrates the efficiency improvements achieved by our approach through a series of experimental results. Finally, conclusion are drawn in Sec. 5.

2. INTRA PREDICTION BY TEMPLATE MATCHING

TMP [3] exploits the self-similarity between different regions within a picture. It is a backward-adaptive prediction technique that derives to the decoder part of the computational effort in order to improve compression efficiency. Encoder and decoder are required to perform the same operations for prediction.

As described in [3], every 4 × 4 block is predicted by means of four 2 × 2 TMP sub-blocks. This implies the use of 4 × 4 blocks for template matching with an overlap of the 75%. The use of 4 × 4 blocks has shown to be a good trade-off between prediction accuracy and matching discrimination. Furthermore, since no overhead information is required, multiple TMP can be fused by means of a weighted superposition of templates in order to form a final improved prediction; reducing in this way the prediction error [5]. In TMP, the template of a block comprises its causal (left and above) neighboring pixels as shown in Fig. 2. The search region is defined, by a large region including only reconstructed pixels available at the decoder. The candidate/s for prediction can be any block in the search region with the same size as the predicting one. The best candidate/s for prediction are those whose associated template most closely resembles the template of the block to predict. Typically the similarity is measured by means of the SAD.

In some cases, illumination variation through the picture may impair SAD based TMP from properly capturing the local characteristics of illumination together with the structural properties of the signal. Fig. 3 clearly shows this situation by means of a toy example. Looking at it, one clearly sees that most similar blocks concentrate along edges. However, a common TMP scheme such as those in [5, 3] will often incur into template selection errors and/or selected templates with an illumination mismatch. This results on a sub-optimal prediction synthesis, leading to a larger residual and an increase of the coding cost. Such illumination variations are often encountered in natural video sequences too.

In the following sections, we show that such mismatch can be adaptively compensated during TMP generation, leading to an improved coding performance for a large range of natural video sequences.

3. ADAPTIVE ILLUMINATION COMPENSATION FOR TEMPLATE MATCHING

3.1. Problem Formulation

Non-uniform scene illumination can result in contrast and/or brightness disparities through a picture. Based on previous works for inter

weighted prediction [6], we thus consider the following linear model to compensate the local illumination mismatch of templates:

\[ \hat{P} = aP + b, \quad (1) \]

where \( P \) and \( \hat{P} \) are, respectively, the reference patch and the illumination compensated signals; \( a \) and \( b \) are factors that jointly account for the contrast and illumination compensation disparity. For every selected template, parameters \( a \) and \( b \) are jointly computed during the search of every template candidate.

Considering Fig. 2, let \( P_i \) and \( X_o \) be the \( i \)th (s.t. \( i \neq 0 \)) reference prediction patch and its associated template respectively. Based on (1), let \( \hat{P}_i \) and \( \hat{X}_b \) be their corresponding illumination compensated versions. Then, taking \( Y \) and \( X_o \) as the current block under prediction and its associated decoded template, we can formulate the distance measure used to select template candidates for TMP as:

\[ \text{Dist} = \| X_o - \hat{X}_b \| = \| X_o - (aX_i + b) \|. \quad (2) \]

Since \( X_o \) and \( X_i \) are reconstructed data known by both encoder and decoder, we can find \( a \) and \( b \) for every tested template by minimizing (2). This can be done by solving the following Least Squares problem.

Rearrange \( X_0 \) and \( X_i \) into two column vectors \( X_0 \) and \( X_i \). Define \( v = [a, b]^T \) as the compensation parameter vector. Then, we can formulate the problem as:

\[ v = \arg \min_v \| X_0 - [X_i, 1]v \|, \quad (3) \]

where \( v \) is an all one column vector with the same length as \( X_i \) and \( X_0 \). For any template containing a number of pixels larger than two, it exist the close form solution of (3), which is:

\[ v = \left(C^T C\right)^{-1} C^T X_0, \quad (4) \]

where \( C = [X_i, 1] \) is an \( N \times 2 \) matrix.

The best candidate templates for TMP are those minimizing (2) subject to (3) and (4). Once the best \( N \) template candidates have been selected, their associate illumination compensated predictions are generated by means of (1). These can then be combined as a weighted superposition [5] in order to generate the final block prediction. In here, each prediction weight can be defined as a scalar proportional to the inverse of its corresponding IC SAD, i.e. (2).

3.2. Algorithm Simplification and Implementation

Although the scheme described above is optimal in order to compensate for both contrast and brightness variations, we must take into account the complexity impact on practical applications of massive
computation of matrix inversions (i.e. (4)) Taking into account the fact that basic prediction blocks in our TMP scheme are only $2 \times 2$, the scaling factor will not be needed in many cases. This means that we can assume parameter $a = 1$, being able to simplify (1) into an offset-only one. The parameter $b$ can be then estimated as:

$$b = E\left[X_0\right] - E\left[X_i\right],$$

and distance in (2) turns into:

$$\text{Dist} = \|X_0 - (X_1 + b)\|$$

$$= \|(X_0 - E\left[X_0\right]) - (X_i - E\left[X_i\right])\|,$$

where $E\left[\cdot\right]$ is the expectation operation. This simplification implies that instead of calculating a $2 \times 2$ matrix inversion for each prediction search, we just need to do it based on a Mean Removed Distance as shown in (6).

The proposed illumination compensation scheme is signaled by an illumination compensation (IC) flag at the macro-block(MB) level, i.e. we define an illumination compensation (IC) flag to signal whether a MB is encoded using IC or not. Once the IC flag is on, a block inside the MB will be coded with illumination compensation if it is a TMP block. Although defining flags for each $4 \times 4$ or $8 \times 8$ block can give better prediction, the increased cost may not be worth. Based on our experimental results, MB level flag is a good tradeoff between the prediction performance and overhead, which works well for most cases.

As shown in [5] and as discussed above, selecting multiple predictions during TMP gives significant compression gain. We implement a multiple template matching approach which selects eight predictions per predicted $2 \times 2$ block. Depending whether the full linear model or the offset-only compensations are used, their respective template selection distance is used for multiple template matching weighting: i.e. (2) or (6). We add our illumination compensated TMP as a new prediction mode to the existing H.264/AVC intra coding modes [9].

In order to asses the impact of the offset-only simplification, tests on the linear model and offset-only model are provided in Sec. 4 for comparison.

## 4. EXPERIMENTAL RESULTS

Our adaptive illumination compensated TMP algorithm has been implemented on JSMV 6 reference software [9]. Three sets of test data are involved in our tests. The first one is the image shown in Fig.3, which is a toy example reproducing an example of correlated structures with non-uniform illumination. The second set is composed by four ten-frame sequences which are cut from four different real movie trailers. Scene 1 shows a building that is lighted by its two wall-mounted lamps. Scene 2 is an indoor shot, in which the sunlight comes into the room from a window and all furniture and actors are lighted non-uniformly. In Scene 3, a table lamp’s light forms an non-uniform illumination scene in a bedroom. Scene 4 shows a house’s wall which is non-uniformly lighted by a outdoor lamp and the lights coming out from a house window. Besides these two test sets which have obvious non-uniform illumination content, we also test our proposed approach on twelve standard test sequences (using 30 frames per sequence), which is our third set of test sequences. The sequences are selected according to common test condition in VCEG [10]. Although some of them (like Crew) include obvious temporal non-uniform illumination, most of them do not include obvious spatial non-uniform illumination.

We compare the average BD rate improvement [11] (average efficiency improvement computed out of a set of 4 R-D quality points) between H.264Intra, H.264Intra+TMP with and without illumination compensation (including with and without offset-only simplification). For this, we test the R-D quality points proposed in [10] (i.e. $QP = 22, 27, 32, 37$). All the frames in each sequence are coded as Intra frame. We use the CA VLC entropy coding method, RD optimization is on, 8 $\times$ 8 transform is on, search range is 64 for TMP, and a single slice per picture. Since our tests do not consider inter prediction, we turned the loop filter off in order to clearly see the TMP effect without any further influence.

Fig. 4 summarizes the BD rate results of the sequences in test set 1 and 2 compared to H.264 Intra. For the toy example frame which has very obvious non-uniform illumination, the BD rate improvement reach more than $19\%$. More than $9\%$ BD rate in average is obtained for the non-standard test sequences. It is obvious that there are more than $4\%$ thanks to our adaptive illumination compensation in average.

Looking to Fig. 5 one can see how the proposed illumination compensation approach can also provide improvements in the set of standard test sequences. In particular Foreman QCIF performance improves more than $16\%$ in average over H.264 intra prediction. Fig. 6 and Fig. 7 illustrate the R-D performance of the Toy Example and Scene 2 respectively. As shown, the proposed approach can
In this paper, we propose a template matching approach with adaptive illumination compensation. This is intended to handle the non-uniform illumination in intra TMP. The proposed illumination compensation technique is based on a linear compensation model with a scale and an offset parameters to jointly compensate local contrast and brightness. The local characteristics of illumination can be adapted by calculating the model parameters based on the template data, which is available at both encoder and decoder. The proposed technique does not require additional information overhead for model parameters. A large set of experiments show that the proposed technique can obtain significant gains compared to H.264/AVC Intra coding. Average gains over 9% BD rate are achieved for sequences with clear illumination variations, and over 5% for standard test sequences which have no obvious illumination variation pattern.

5. CONCLUSIONS

6. REFERENCES