An Optimized Template Matching Approach to Intra Coding in Video/Image Compression

Hui Su, Jingning Han, and Yaowu Xu

Chrome Media, Google Inc., 1950 Charleston Road, Mountain View, CA 94043

ABSTRACT

The template matching prediction is an established approach to intra-frame coding that makes use of previously coded pixels in the same frame for reference. It compares the previously reconstructed upper and left boundaries in searching from the reference area the best matched block for prediction, and hence eliminates the need of sending additional information to reproduce the same prediction at decoder. In viewing the image signal as an auto-regressive model, this work is premised on the fact that pixels closer to the known block boundary are better predicted than those far apart. It significantly extends the scope of the template matching approach, which is typically followed by a conventional discrete cosine transform (DCT) for the prediction residuals, by employing an asymmetric discrete sine transform (ADST), whose basis functions vanish at the prediction boundary and reach maximum magnitude at far end, to fully exploit statistics of the residual signals. It was experimentally shown that the proposed scheme provides substantial coding performance gains on top of the conventional template matching method over the baseline.

Keywords: Template matching, Intra prediction, Transform coding, Asymmetric discrete sine transform

1. INTRODUCTION

Intra-frame coding is a key component in video/image compression system. It predicts from previously reconstructed neighboring pixels to largely remove spatial redundancies. A codec typically allows various prediction directions[1–3] , and the encoder selects the one that best describes the texture patterns (and hence rendering minimal rate-distortion cost) for block coding. Such boundary extrapolation based prediction is efficient when the image signals are well modeled by a first-order Markovian process. In practice, however, image signals might contain certain complicated patterns repeatedly appearing, which the boundary prediction approach can not effectively capture. This motivates the initial block matching prediction, that searches in the previously reconstructed frame area for reference, as an additional mode.[4] A displacement vector per block is hence needed to inform decoder to reproduce the prediction, akin the motion vector for inter-frame motion compensation. To overcome such overhead cost that diminishes the performance gains, a template matching prediction (TMP) approach was developed[5] that employs the available neighboring pixels of a block as a template, measures the template similarity between the block of interest and the candidate references, and chooses the most "similar" one as the prediction. Clearly the decoder is able to repeat the same process without recourse to further information, which further allows the TMP to operate in smaller block size for more precise prediction at no additional cost. A conventional 2D-DCT is then applied to the prediction residuals, followed by quantization and entropy coding, to encode the block. Certain coding performance gains were obtained by integrating the TMP in a regular intra coder.

In viewing the image signals as an auto-regressive process, pixels close to the block boundaries are more correlated to the template pixels, and hence are better predicted by the matched reference, than those sitting at far end. Therefore, the residuals ought to exhibit smaller variance at the known boundaries and gradually increased energy to the opposite end, which makes the efficacy of the use of DCT questionable due to the fact that its basis functions get to the maximal magnitude at both ends and are agnostic to the statistical characteristics of the residuals. This work addresses this issue by incorporating the ADST,[6,7] whose basis functions possess the desired asymmetric properties, as an alternative to the TMP residuals for optimal coding performance. A complementary similarity measurement based on weighted template matching, in recognition of the statistical

E-mails: {huisu, jingning, yaowu}@google.com.

Visual Information Processing and Communication V, edited by Amir Said, Onur G. Guleryuz,
Robert L. Stevenson, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 9029, 902904
© 2014 SPIE-IS&T · CCC code: 0277-786X/14/$18 · doi: 10.1117/12.2040890

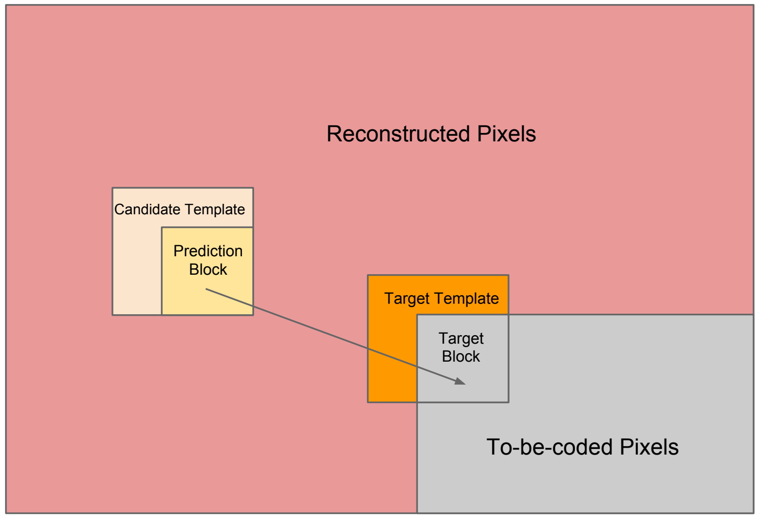

Figure 1. Template matching intra prediction.

variations across the block, was also proposed to improve the search quality. The scheme was implemented in the VP9 framework, in conjunction with other boundary prediction based intra coding modes. Experiments demonstrated remarkable performance advantages over conventional TMP as well as the baseline codec.

The rest of the paper is organized as follows: Sec. 2 presents a brief review on the template matching approach. In Sec. 3, we describe the proposed techniques in details. Experimenting results are presented in Sec. 4, and Sec. 5 concludes the paper.

2. REVISITING TEMPLATE MATCHING PREDICTION

We provide a brief review on the TMP approach[5] in this section. As shown in Fig. 1, the TMP employs the pixels in the adjacent upper rows and left columns of a block as its template. Every template in the reconstructed area of the frame is considered as a reference template, and the template of the block to be encoded is the target template. The similarity between the target template and the reference templates is then evaluated in terms of sum of absolute/squared difference. The encoder selects amongst the reference templates the one that best resembles the target template as the candidate template, and the block corresponding to this candidate template is used as the prediction for the target block. Since it only involves comparing reconstructed pixels, the same operations can be repeated at the decoder side without any additional side information sent, resulting in higher compression efficiency than the direct block matching approach.[4] As a consequence, the decoding process gets more computationally loaded.

The TMP approach was shown to be particularly efficient in the scenarios where certain complicated texture patterns, that cannot be captured by the conventional directional intra prediction modes, appear repeatedly in the image/frame. Recent research efforts have been devoted to further improve the TMP scheme, including combining multiple candidates with top similarity scores,[8] using hybrid TMP and block matching (with displacement vector sent explicitly),[9] etc. This work is focused on optimizing the original TMP approach by observing and exploiting the statistical property of the TMP residual signals. It is noteworthy that the proposed principles are generally applicable to other advanced variants as well.

3. PROPOSED TECHNIQUES

We view the image signals as an auto-regressive model, which implies that two nearby pixels are more correlated than those far apart. Since the template of a matched reference block closely resembles that of the block of

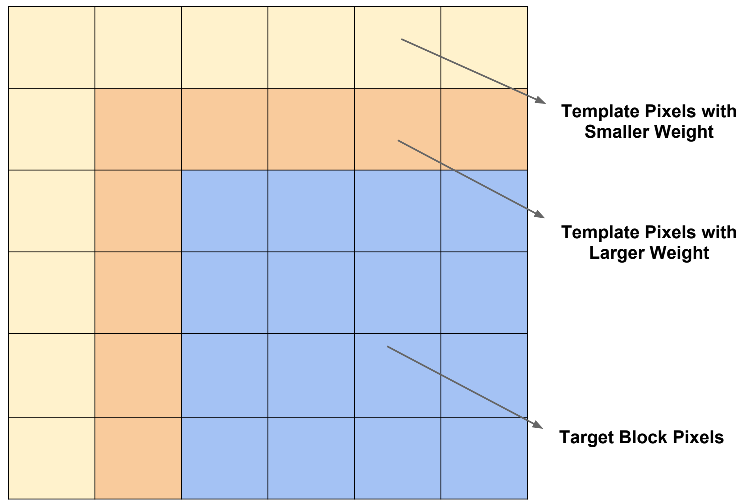

Figure 2. The proposed weighted template matching scheme. The pixels that are closer to the target block are assigned with larger weight.

interest, the pixels sitting close to the known boundaries of the two blocks are element-wisely more correlated, than those at the opposite end. Hence the pixels near the top/left boundaries are better predicted by the matched reference block, which translates into a key observation that the variance of prediction residuals tends to vanish at the known boundaries and gradually increase towards the far end. This inspires that unlike the discrete cosine transform (DCT) whose basis functions get maximum magnitude at both ends, the (near) optimal spatial transform for the TMP residuals should possess such asymmetric properties. We hence propose to employ the asymmetric discrete sine transform (ADST)[6,7] for transform coding of the TMP residuals. A complementary matching approach that expands the template to multiple boundary rows and columns, and uses a weighted sum of difference measurement is first developed for more precise referencing. A statistical study of the TMP residuals, followed by the detailed discussion of ADST will be provided next.

3.1 Weighted Template Matching

In order to obtain reliable template matching, it is reasonable to define multiple layers of boundary pixels as the template of a block. In our study, we have observed that the prediction accuracy can be improved when the number of rows and columns in the template increases. However, it is not wise to adopt too many layers, as the gain in matching accuracy becomes saturated and the computational complexity explodes. In our implementation, the template consists of the pixels in the 2 rows and 2 columns above and to the left of the given block, which gives a good tradeoff between accuracy and computation complexity.

The similarity between the target templates and reference templates can be measured by the sum of absolute difference (SAD). Along the line of recognizing the variations in statistics, the template pixels closer to the block are highly correlated to the block content, and hence should be more weighted in the SAD calculation than those distant ones. This idea is illustrated in Fig. 2. A weight ratio of 3:2 for the inner row/column versus the outer row/column is used in this work.

3.2 Spatial Transformation

In video/image compression, the prediction residuals are typically processed via transformation to further remove the remaining spatial redundancy, before the quantization and entropy coding modules. The Karhunen-Loeve transform (KLT) is considered as the optimal spatial transform in terms of energy compaction. However, KLT is rarely used in practical coding system due to its high computation complexity. The DCT has long been a popular substitute due to its good tradeoff between energy compaction and complexity. The basis functions of the DCT are as follows:

$$[T_C]_{j,i} = \left(\alpha \ cos\frac{\pi(j-1)(2i-1)}{2N}\right), \tag{1}$$

where N is the block size, $i, j \in \{1, 2, \cdots, N\}$ denote the space and frequency indexes, respectively; and

$$\alpha = \begin{cases} \sqrt{\frac{1}{N}}, & \text{if } j = 1 \\ \sqrt{\frac{2}{N}}, & \text{otherwise} \end{cases}$$

It is easy to see that the basis functions of the DCT achieve their maximum energy at both ends (i.e., $i = 1$ or $i = N$).

Assuming the template of a matched reference block closely approximates that of the block of interest, it is highly likely that pixels close to these known boundaries are also well predicted, while those distant pixels are less correlated, which results in a relatively higher residual variance. This postulation is verified by the following experimental study. We collected the absolute values of the TMP prediction residues element-wisely over 8000 blocks (of dimension 4 × 4) from the foreman sequence, and the average of the residue signal at each pixel location was calculated, as shown below:

$$\begin{pmatrix} 4.05 & 4.37 & 4.51 & 5.13 \\ 4.72 & 5.48 & 5.50 & 6.60 \\ 5.04 & 5.95 & 6.12 & 7.32 \\ 5.50 & 6.28 & 6.97 & 8.20 \end{pmatrix}$$

As can be seen from the matrix, the variance of the prediction residue signal indeed increases along both the horizontal and vertical directions.

As abovementioned, the basis functions of the conventional DCT achieve their maximum energy at both ends and are therefore agnostic to the statistical patterns of the prediction residuals. As an alternative, the ADST[6,7] has basis functions of form:

$$[T_S]_{j,i} = \left(\frac{2}{\sqrt{2N+1}} sin\frac{(2j-1)i\pi}{2N+1}\right), \tag{2}$$

where N is the block size, $i, j \in \{1, 2, \cdots, N\}$ denote the space and frequency indexes, respectively. It is shown[6,7] that the ADST is a better approximation of the optimal KLT than the DCT when the partial boundary information is available. Clearly, the basis functions of ADST vanishes at the known prediction boundary ($i = 1$) and maximizes at the far end ($i = N$), and therefore matches well with the statistical patterns of the TMP residuals. We hence propose to employ the ADST as the spatial transform for the TMP residuals. It is experimentally shown in the next section that the use of ADST provides substantial performance improvement over the TMP followed by the conventional DCT.

4. EXPERIMENT RESULTS

The proposed scheme was tested in the VP9 framework.[1] We verified its efficacy in a relatively simplified setting, where the block size was fixed as 8 × 8. There are 10 intra prediction modes in VP9, including vertical prediction, horizontal prediction, 8 angular prediction modes, and a "true motion" mode that utilizes the left, above and corner pixels simultaneously. The TMP scheme was implemented as an additional mode to the 10 existing ones. The selection among the 11 modes is based on rate-distortion optimization. In the TMP mode, the 8 × 8 block is further partitioned into four 4 × 4 blocks, each of which is predicted via template matching, followed by the 2D-ADST transform, quantization, and reconstruction, in a raster scan order. The template consists of pixels from 2 rows and 2 columns above and to the left of the given block. For the weighted template matching, we use a weight ratio 3:2 for the inner row/column versus the outer row/column, as shown in Fig. 2.

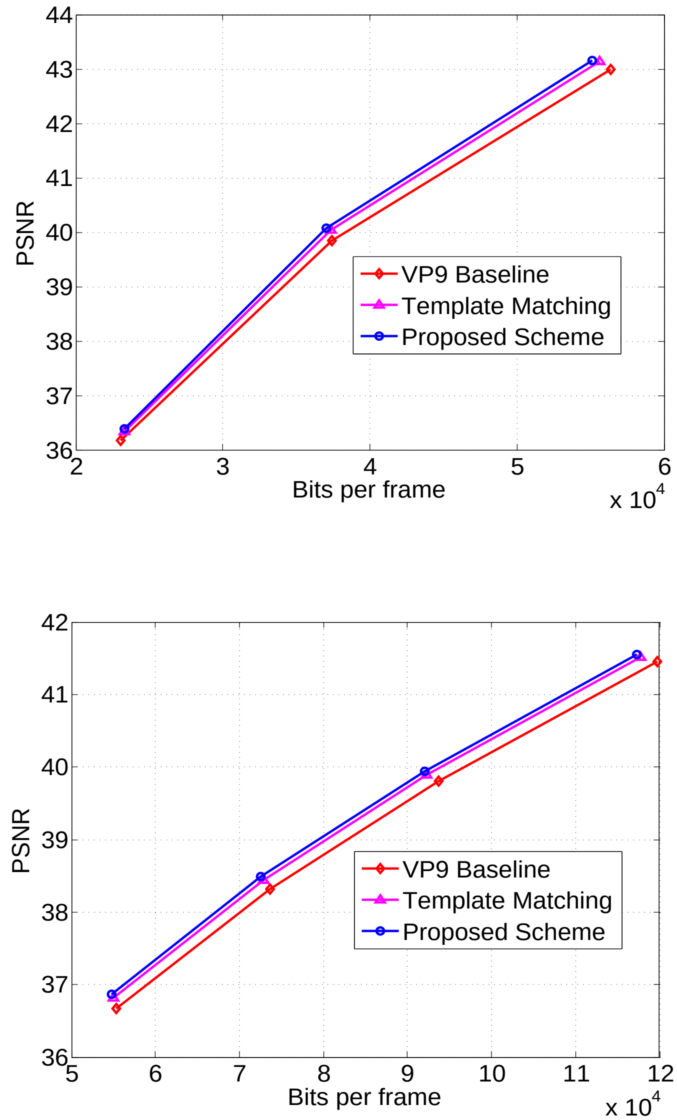

Figure 3. Rate-Distortion curves of the Ice (upper) and Foreman (lower) test sequences.

Several test video clips were used to compare the coding efficiency, including the Ice, Foreman, and Carphone sequence. For every test sequence, the first 75 frames were coded as key-frame (i.e., all blocks were coded in intra modes), at various bit-rates. The coding performance gains of the conventional TMP and the proposed method over the reference codec, measured by the Bjontegaard metric, are shown in Table 1. Clearly, the proposed approach that optimizes the transformation for prediction residual significantly improves the performance of TMP, and both outperform the reference VP9 baseline. The rate-distortion curves of the ice and foreman sequence are also provided in Fig. 3. It can be seen from the figure that the proposed techniques boost the coding efficiency of the conventional TMP consistently.

5. CONCLUSIONS AND FUTURE WORK

This work proposed a novel approach that incorporated the ADST for TMP prediction residuals as an additional mode for intra-frame coding. A complementary template matching method along the lines of recognizing the

Table 1. Coding performance gains over VP9 baseline in terms of bit-rate reduction percentage.

Sequences	Conventional TMP	Proposed Method
Ice	2.89	3.78
Foreman	2.88	3.33
Carphone	1.05	1.35

statistical variations across block was also provided for more precise reference search. The scheme implemented in the VP9 framework demonstrated substantial performance improvements over the conventional TMP as well as the reference codec.

The TMP approach can also be applied to inter-frame prediction.[10] The template of a block is defined as the pixels in the adjacent upper rows and left columns, in the same way as in the case of intra prediction. The optimal template which is best matched to that of the block of interest is found in a previously encoded reference frame, and the block to be encoded is filled in by copying the block corresponding the optimal template. By the same principles in this work, the residue signal of the template matching inter prediction should also present asymmetric statistical property across the block. We thus expect the ADST to be more efficient than the conventional DCT for the transform coding of the template matching inter prediction, and are currently working along this direction.

REFERENCES

[1] *VP9 Video Codec* , http://www.webmproject.org/vp9/.

[2] Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A., "Overview of the H.264/avc video coding standard," *IEEE Trans. Circuits and Systems for Video Technology* **13**, 560–576 (July 2003).

[3] Sullivan, G., J. Ohm, W. H., and Wiegand, T., "Overview of the high effciency video coding (HEVC) standard," *IEEE Trans. Circuits and Systems for Video Technology* **22**, 1649–1668 (Dec. 2012).

[4] Yu, S. and Chrysafis, C., "New intra prediction using intra-macroblock motion compensation," *Tech. Rep. JVT-C151* (2002).

[5] Tan, T., Boon, C., and Suzuki, Y., "Intra prediction by template matching," *IEEE Proc. ICIP* , 1693–1696 (2006).

[6] Han, J., Saxena, A., and Rose, K., "Towards jointly optimal spatial prediction and adaptive transform in video/image coding," *IEEE Proc. ICASSP* , 726–729 (2010).

[7] Han, J., Saxena, A., Melkote, V., and Rose, K., "Jointly optimized spatial prediction and block transform for video and image coding," *IEEE Trans. on Image Processing* **21**, 1874–1884 (2012).

[8] Tan, T., Boon, C., and Suzuki, Y., "Intra prediction by averaged template matching predictors," *IEEE Proc. CCNC* (2007).

[9] Cherigui, S., Thoreau, D., Guillotel, P., and Perez, P., "Hybrid template and block matching algorithm for image intra prediction," *IEEE Proc. ICASSP* , 781–784 (2012).

[10] Sugimoto, K., Kobayashi, M., Suzuki, Y., Kato, S., and Boon, C. S., "Inter frame coding with template matching spatio-temporal prediction," *IEEE Proc. ICIP* (2004).