# FAST TEMPLATE MATCHING FOR INTRA PREDICTION

*Gayathri Venugopal, Philipp Merkle, Detlev Marpe and Thomas Wiegand*

Video Coding & Analytics Department, Fraunhofer Heinrich Hertz Institute (HHI), Berlin 10587, Germany

## ABSTRACT

Intra prediction is an important step in video coding, where spatial redundancy is exploited for coding efficiency. The state-of-the-art H.265/High Efficiency Video Coding (HEVC) standard provides 35 intra prediction modes including 33 angular modes and a DC as well as a Planar mode. Although these intra prediction methods are quite efficient, there is still room for improvements, especially when complex image structures are involved. Template matching was proposed previously for H.264/Advanced Video Coding (AVC) with up to 11% BD-rate gain. However substantial decoder complexity increase from the implied search algorithm makes it less attractive for practical applications. In this paper, we propose a fast template matching intra mode for HEVC. The proposed mode is scalable in the sense of providing different trade-offs between decoder complexity and BD-rate gains. For a specific choice of parameters, fast template matching mode can achieve an average BD-rate gain of 1.15% with a maximum BD-rate gain of up to 4.6% at a 33% increase in decoder run-time. Overall, fast template matching can provide the same coding gain as conventional template matching approaches, but with significantly less decoder complexity.

***Index Terms*—** Intra prediction, template matching, High Efficiency Video Coding

## 1. INTRODUCTION

In a block-based video coding standard, like H.265/High Efficiency Video Voding (HEVC) or H.264/Advanced Video Coding (AVC), an input picture is first partitioned into blocks. Each block is then predicted by using either intra or inter prediction. While the former prediction method uses only decoded samples within the same picture as a reference, the latter uses displaced blocks of already decoded pictures [1]. Once the predicted blocks are obtained, they are used to calculate residual blocks which are further processed before encoding. Information related to the tools and techniques applied at the encoder are to be send to the decoder for reconstruction of the images.

HEVC provides 35 intra prediction modes [2]. The encoder typically chooses the best intra mode in the sense of minimizing a given cost function and that mode is signaled in the bitstream to the decoder. The commonly used cost function $J$ is defined as

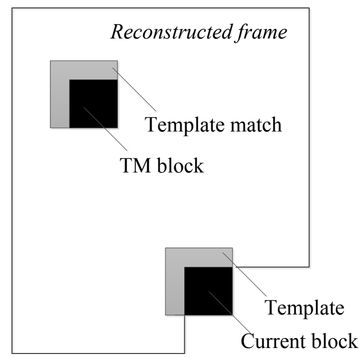



**Fig. 1**: Template matching intra prediction

$$J = D + \lambda R, \tag{1}$$

where $D$ is the distortion between the original and predicted blocks, $R$ is the number of bits associated with the intra mode and $\lambda$ is the Lagrange parameter that determines the trade-off between $D$ and $R$ [1].

The rest of the paper is organized as follows. In section 2 conventional template matching (TM) is introduced. The research methodology and the proposed fast TM mode is explained in section 3. Section 4 contains experimental conditions and results. Finally, section 5 concludes the paper.

## 2. CONVENTIONAL TEMPLATE MATCHING

Template matching (TM) is a texture synthesis technique used in digital image processing, which can be applied for intra prediction as well. A patch of already decoded samples present above and left of the current block is called the template (Fig.1). TM finds the best match for the template in the reconstructed frame by minimizing the error between the template and its match, usually measured as sum of squared differences (SSD). Finally, in TM-based prediction, the TM block associated with the error minimizing template match is used as the prediction of the current block. This TM-based prediction approach does not require any side information for generating the corresponding prediction signal at the decoder, because the same search process is performed there as well. On the other hand, the whole search process for the best template match leads to high computational complexity. Usually,

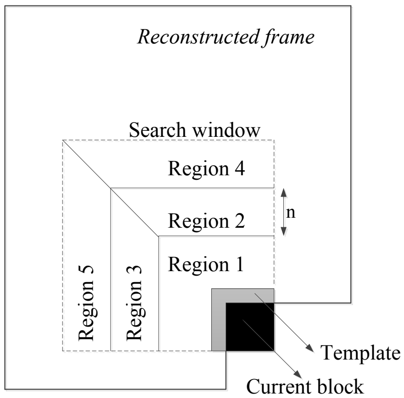



**Fig. 2**: Five regions of the proposed TM intra mode

restricting the search process inside a window is considered for keeping the complexity increase manageable[3, 4].

Intra prediction using TM was first proposed in [3] for H.264/AVC with up to 11% coding efficiency gains. In [4] from the same authors, TM with more than one predictor was proposed. Other proposals related to TM intra prediction can be found in the literature [5, 6, 7, 8, 9, 10, 11]. All the works mentioned above concentrate more on the coding-efficiency aspect of TM without considering much on the implied decoder complexity of the approach.

## 3. PROPOSED FAST TEMPLATE MATCHING

As mentioned previously, typical TM intra mode results in high decoder complexity, as a consequence of the fact that no side information is being sent. Our research targets to reduce the decoder complexity by making small compromises on the coding efficiency. Initial investigations on the impact of different parameters for TM-based intra prediction, like the template size or choice of multiple predictors, were conducted with HEVC as the anchor. Based on the outcome of these experiments, two-sample wide rotated-L shaped template and the superposition of three predictors were decided for the proposed mode, which operates as follows: Consider an $N \times N$ block to be predicted. In the reconstructed frame, $(N + 2) \times 2$ samples above and $2 \times N$ samples left of the block is taken as the template. The proposed method searches for the best template matches inside a given search window, whose size is adapted to the frame size of the input video sequence. This window is further divided into five regions (Fig.2) and three best template matches from each region are found by minimizing SSD. The averaged superposed signal of the respective TM blocks is used as the prediction of the current block (Fig.3). The rate-distortion optimization algorithm of the HEVC encoder according to (1) is utilized for deciding which region gives the best prediction.

The information related to the chosen region is signaled in the bitstream. The decoder searches for the best matches only

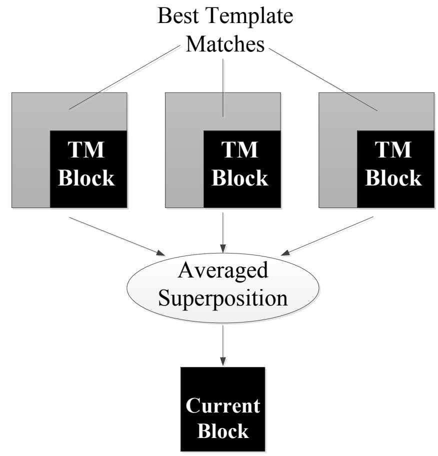



**Fig. 3**: Superposition of three predictors in the fast TM mode

in that signaled region and uses the signal of the averaged superposed blocks that correspond to the best template matches for the final prediction. The aforementioned intra mode is added to the encoder candidate list of luma prediction modes for all available intra block sizes along with its 35 other intra modes.

## 4. EXPERIMENTAL RESULTS

We have used the HEVC test model reference software (HM version 16.6) and the common test conditions [12] for our experiments. Experiments were restricted to All Intra (AI) main configuration. Subsection 4.1 to 4.4 are related to the optimization approaches adopted for the new mode and subsection 4.5 contains results for the proposed mode.

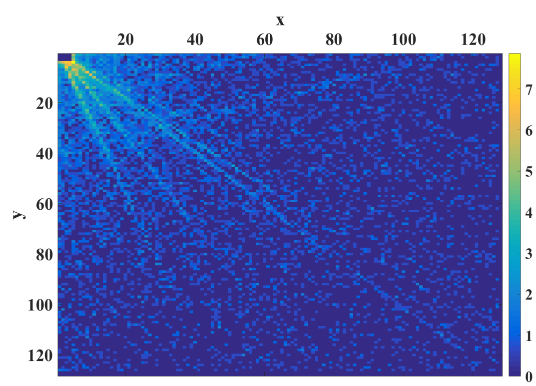



**Fig. 4**: Histogram of the number of TM blocks with respect to the current block (log scale) inside a $128 \times 128$ window for BasketBallDrill of QP = 22

### 4.1. Position of TM block with respect to current block

In order to get an idea of how to reduce the complexity of the search algorithm, we investigated the position of the TM block with respect to the location of the current block. Our study revealed that the TM block is more often present in the neighbourhood of the current block (Fig.4). This clearly justifies the use of windowing approach for restricting the search process in our proposed mode.

### 4.2. Search window size

An investigation into the effect of the search window size without adapting to the frame size was carried out. Window sizes $M$ = 8, 16, 24, 32, 64, 128 were considered for this. The general tendency from the research result is that the bigger the window, the better the performance (Fig.5). However, there are three main observations to be noted here. First, the gain saturates after a particular value of $M$ and second, this value of $M$ varies with different classes (each class has a different resolution, see Table 3). Third, large window sizes result in high complexity as in Table 1. These remarks lead to the idea of an adaptive window for searching in the new mode, where the window size varies with the frame size of the sequence.

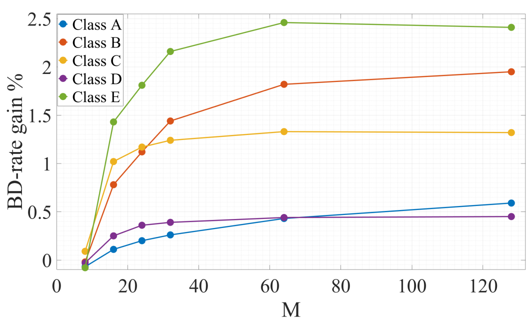



**Fig. 5**: Search window size $M$ vs. coding efficiency

| $M$ | **8** | **16** | **24** | **32** | **64** | **128** |
|---|---|---|---|---|---|---|
| Decoder | 16 | 33 | 66 | 109 | 350 | 973 |
| Encoder | 17 | 28 | 49 | 76 | 216 | 573 |

**Table 1**: Search window size $M$ vs. run-time increase (%)

### 4.3. Position of second best TM block with respect to first

In the proposed TM mode, the averaged superposition of the three best TM blocks in the sense of SSD minimization is used for prediction. For reducing complexity, it was important to understand the position of the second and third best TM blocks with respect to the first one (in terms of SSD minimization). We wanted to know whether the three blocks are spread out or near to each other. Our research result show that they are more often close to one another (Fig.6). This observation motivated our idea of dividing the search window into different regions. Based on our experimental results, an adaptive search window is proposed for our new mode, where its size is varied depending on the frame size. The window size is $M \times M$ where $M = n \times 3$ (Fig.2) and $n$ is calculated according to Table 2. It should be noted that the proposed adaptive search window is applied only to smaller blocks.

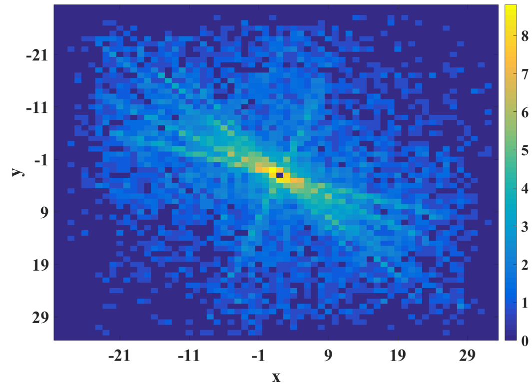



**Fig. 6**: Histogram of the number of second best TM blocks with respect to the first block (log scale), for BasketBallDrill of QP = 22

| **Block size, $N$** | **framewidth, $w$** | $n$ |
|---|---|---|
| 4, 8 | $0 \leq w \leq 832$ | $A$ |
| | $832 < w < 1280$ | $2 \times A$ |
| | $1280 \leq w < 2560$ | $3 \times A$ |
| | otherwise | $4 \times A$ |
| 16, 32 | always | $2 \times A$ |

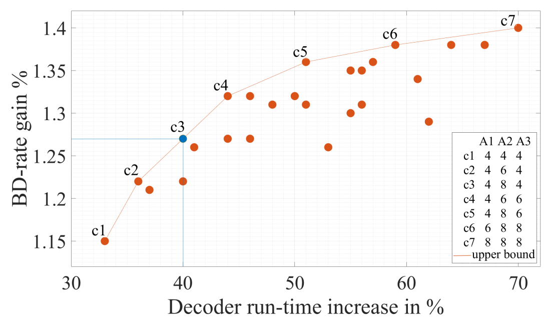
**Table 2**: Adaptive search window for Fast TM

### 4.4. Further investigation on value of $A$

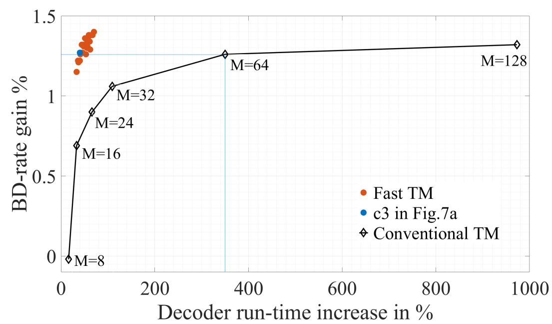
The parameter $A$ determines the size of the search window as shown in Table 2, which directly affects the run-time of the proposed algorithm. Consequently, the value of $A$ is crucial. We consider the value of $A$ for Region1 to be $A1$, for Region2 and 3 to be $A2$ and, for the rest as $A3$. Similarly, $n1$ is related to $A1$, $n2$ to $A2$ and $n3$ to $A3$. The search window is adapted according to Table 2. We carried out tests for different combinations of $A1$, $A2$ and $A3$ with $A$ = 4, 6, 8. Our experimental results indicate that further coding efficiency can be achieved by unequal values of $A$ (Fig.7a).

### 4.5. Fast TM Mode for HEVC

The proposed mode with $A = 4$ can achieve an average BD-rate gain of 1.15% for classes A to E with 133% decoder complexity (Table 3, $c1$ in Fig.7a). The fast TM mode has much lesser decoder run-time increase when compared to conventional TM (Fig.7a,7b). While conventional TM achieves 1.26% coding gain with 350% increase in decoder run-time for $M$ = 64, fast TM achieves the same gain with just 40% increase in decoder run-time (nearly 9 times less). The results in Fig.8 show that an average area of 10.6% of a frame is predicted by TM mode. The evaluation further indicates that TM mode works best for sequences with homogeneous structures, like BasketBallDrill.




(a)




(b)

**Fig. 7**: (a) Fast TM for different combinations of $A1$, $A2$, $A3$ and (b) Fast TM vs. conventional TM. Highlights the run-time differences of the two TM approaches for the same coding gain.

### 4.6. Proposed mode Signalization

The fast TM mode is indicated by an additional flag relative to the HEVC-compliant signaling of intra prediction modes. If true, one more flag is signalled for the region of search and if false, conventional HEVC signalling is followed. If Region1 is the chosen search area, the second flag is set to true. Otherwise, 2 more bins are used for representing one of the regions from 2 to 5. The Context-Adaptive Binary Arithmetic Coding (CABAC) is applied for signalling [1, 13] the flags and the bins. Two separate context models are used for the two bins.

## 5. CONCLUSION AND FUTURE RESEARCH

We presented a fast template matching (TM) intra mode for HEVC, targeting a better trade-off between coding efficiency and decoder run-time than the conventional TM approaches. The results show that fast TM can achieve the same coding gain with about 9 times less decoder run-time than conventional TM methods. The proposed mode searches for the best template match inside a window, whose size is adapted to the frame size of the video sequence. Further, the search window is partitioned into regions, such that the proposed method can be tuned to provide varying combinations of BD-rate gain and decoder run-time, depending on the needs of the chosen application. Our research underlines the fact that a good prediction from a local search often provides a much better trade-off than an optimal prediction from a highly time consuming global search. The experimental results indicate that for a specific choice of parameters, the proposed TM mode can provide an average BD-rate gain of 1.15% for a representative test set of video sequences with 33% increase in decoder run-time. Additional reduction in computational complexity could be achieved by terminating the search algorithm after reaching a minimum threshold. This will be a subject of future research.

| **Test Sequences** | | **BD Rate (Y)** |
|---|---|---|
| Class A: 2560 x 1600 | Traffic | -0.49% |
| | PeopleOnStreet | -0.84% |
| | Nebuta Festival | -0.10% |
| | SteamLocomotive | -0.16% |
| Class B: 1920 x 1080 | Kimono | -0.63% |
| | ParkScene | -0.45% |
| | Cactus | -2.36% |
| | BQTerrace | -2.43% |
| | BasketBallDrive | -2.15% |
| Class C: 832 x 480 | RaceHorsesC | -0.12% |
| | BQMall | -0.72% |
| | PartyScene | -0.54% |
| | BasketBallDrill | -3.23% |
| Class D: 416 x 240 | RaceHorses | 0.01% |
| | BQSquare | -0.80% |
| | BlowingBubbles | -0.18% |
| | BasketBallPass | -0.30% |
| Class E: 1280 x 720 | FourPeople | -1.28% |
| | Johnny | -4.60% |
| | KristenAndSara | -1.64% |
| **Average** | | **-1.15%** |
| **Encoder run-time** | | **180%** |
| **Decoder run-time** | | **133%** |

**Table 3**: Results of the Fast TM mode for HEVC

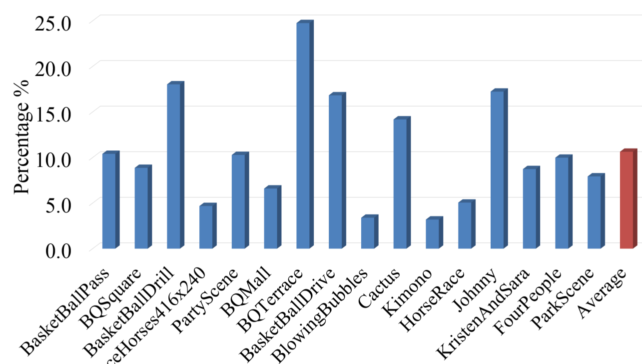



**Fig. 8**: Average percentage of area in a frame predicted by TM mode for different sequences

## 6. ACKNOWLEDGEMENT

This research was conducted under the PROVISION ITN, which was funded from the European Unions Seventh Framework Programme under grant agreement number 608231.

## 7. REFERENCES


[1] Vivienne Sze, Madhukar Budagavi, and Gary J. Sullivan, *High Efficiency Video Coding (HEVC) Algorithms and Architectures*, Springer, 2014.

[2] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding HEVC standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, September 2012, vol. 22, pp. 1649 – 1668.

[3] Thiow Keng Tan, Choong Seng Boon, and Yoshinori Suzuki, "Intra prediction by template matching," in *IEEE International Conference on Image Processing (ICIP)*, Altlanta, GA, 2006, pp. 1693 – 1696.

[4] Thiow Keng Tan, Choong Seng Boon, and Yoshinori Suzuki, "Intra prediction by averaged template matching predictors," in *Proc. CCNC 2007*, Las Vegas, NV, USA, 2007, pp. 405–109.

[5] Yi Guo, Ye-Kui Wang, and Houqiang Li, "Priority-based template matching intra prediction," in *IEEE International Conference on Multimedia and Expo 2008*, Hannover, Germany, June 2008, pp. 1117–1120.

[6] Yinfei Zheng, Peng Yin, Oscar Divorra Escoda, Xin Li, and Cristina Gomila, "Intra prediction using template matching with adaptive illumination compensation," in *Proc. IEEE International Conference on Image Processing ICIP2008*, San Diego, CA, October 2008.

[7] Matthieu Moinard, Isabelle Amonou, Pierre Duhamel, and Patrice Brault, "A set of template matching predictors for intra video coding," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2010*, March 2010.

[8] Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel, and Patrick Perez, "Hybrid template and block matching algorithm for image intra prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, March 2012, pp. 781–784.

[9] Tao Zhang, Haoming Chen, Ming-Ting Sun, Debin Zhao, and Wen Gao, "Hybrid angular intra/template matching prediction for hevc intra coding," in *Visual Communications and Image Processing 2015*, Singapore, December 2015, pp. 1–4.

[10] Cuiling Lan, Jizheng Xu, Feng Wu, and Guangming Shi, "Intra frame coding with template matching prediction and adaptive transform," in *IEEE International Conference on Image Processing (ICIP)*, September 2010.

[11] Karam Naser, Vincent Ricordel, and Patrick Le Callet, "Local texture synthesis: A static texture coding algorithm fully compatible with HEVC," in *International Conference on Systems, Signals and Image Processing, IWSSIP*, September 2015.

[12] F.Bossen, "Common test conditions and software reference configurations," in *JCTVC-L1100 of JCT-VC*, Geneva, 2013.

[13] Detlev Marpe, Heiko Schwarz, and Thomas Wiegand, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," in *IEEE Transactions on Circuits and Systems for Video Technology*, July 2003, vol. 13, pp. 620–636.