Intra Prediction by a linear combination of Template Matching predictors

Laurent Guillo, Tangi Poirier, Christine Guillemot

Joint Collaborative Team on Video Coding (JCT-VC)
of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11
2nd Meeting: Geneva, CH, 21-28 July, 2010

Document: JCTVC-B078

Title:	**Intra Prediction by a linear combination of Template Matching predictors**
Status:	Input Document to JCT-VC
Purpose:	Proposal

Author(s) or Contact(s): L. Guillo (IRISA/CNRS) Tel: (+33) (0) 299847100 Email: Laurent.Guillo@inria.fr

T. Poirier (INRIA)

C. Guillemot (INRIA)

INRIA Rennes Bretagne Atlantique

Campus de Beaulieu

35042 Rennes Cedex

France

Source: INRIA.

Abstract

In intra mode, a prediction block is formed by extrapolating pixels neighbouring the current block to be coded. It is then subtracted from the current block prior encoding the resulting residual. When blocks have regular textures, this method is efficient. However, predicting blocks with more complex textures in this way is less adapted. This contribution presents an intra prediction technique using template matching enhanced by a weighting factor computed by a one step matching pursuit approach. This method can be seen as a generalization of the template matching method in which the weighting factor is not “1”. It is also more general in the sense that it can lead to a linear combination of blocks which will best approximate the template, hence the block to be predicted. This only tool has been integrated in KTA2.7 and compared with the original KTA2.7. Classes and videos used are those selected by the Intra AhG. All pictures were encoded as Intra. Improvements related to videos belonging to the classes A, B1 and D are scarcely significant. However, average gains for videos belonging to the classes C, E and B2 are respectively (+0.17dB, -2.56%), (+0.19dB, -3.33%) and (+0.17dB, -3.79%).

1 Introduction

In H264/AVC, the prediction is based on the knowledge of the pixels surrounding the current block to be coded. These reconstructed pixels are extrapolated to form the sample predictor. It is then subtracted from the current block prior encoding the resulting residual. Several kinds of extrapolations of reconstructed pixels related to the block size are possible. The standard specifies a DC and 8 directional modes for 4x4 and 8x8 blocks and only DC, horizontal, vertical and planar for 16x16 blocks [1]. H264/AVC intra coding is efficient when textures of blocks to be predicted are regular, uniform or have directional structures which fit well with the direction of one of the intra modes.

However, this kind of intra prediction is less adapted when blocks are highly textured. Alternative intra prediction methods based on template matching or sparse representation were respectively proposed in [2] and [3] and in [4].

This contribution presents a new extension of template matching taking advantage of a weighting factor computed thanks to a one step matching pursuit approach. It has been implemented for 4x4, 8x8 and 16x16 block sizes. Section 2 describes this new approach and illustrates it with an example. This approach was integrated in KTA 2.7 [5] and the resulting codec was compared with KTA2.7. Section 3 gathers these results when this approach is applied on a subset of videos used for the latest Call for Proposals (CfP) issued by ITU-T SG16 Q.6 (VCEG) and ISO/IEC JTC1/SC29/WG11 (MPEG). Section 4 presents on-going works and future extensions of this method.

2 Intra prediction by weighted template matching

2.1 Description of the algorithm

The method described in this section aims at predicting a block thanks to pieces of picture belonging to the causal neighbourhood, i.e. thanks to blocks which have already been reconstructed. It has been implemented for intra prediction of 4x4, 8x8 and 16x16 blocks. The following explanations are related to 4x4 blocks only. Extension to larger sizes is analogous and is compliant with the prediction order of blocks.

The algorithm relies on a search region adjacent to the current block to be predicted as pictured in Figure 1.

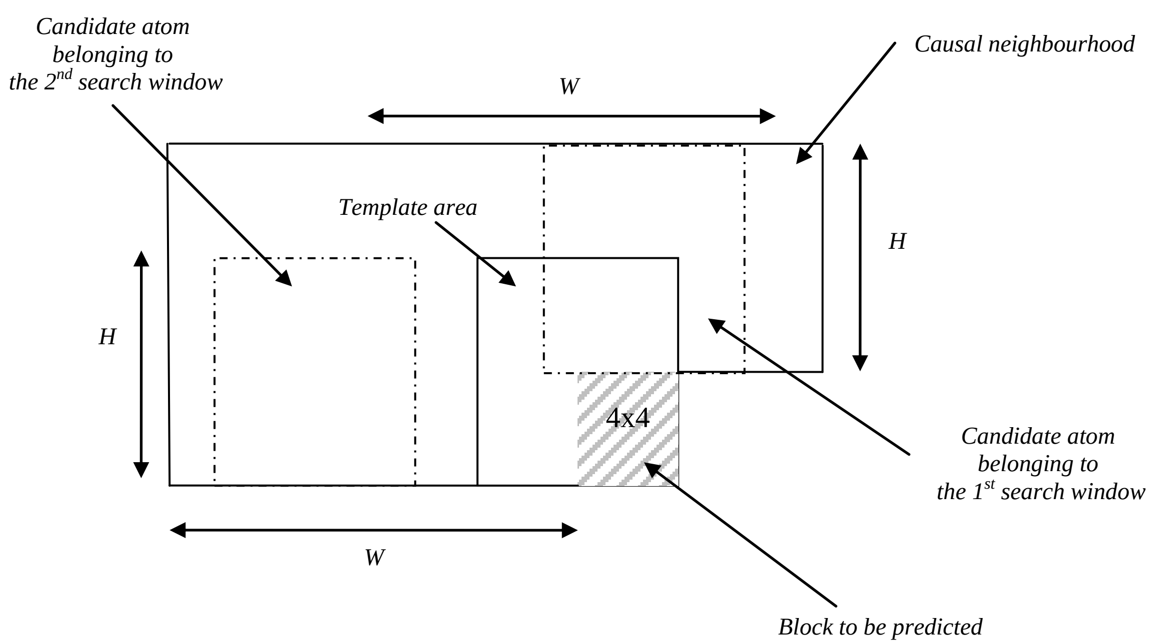

Figure 1 : Definition of search regions from causal neigbourhood

In the previous figure, the block to be predicted is a 4x4 block and is indicated with grey lines. It is surrounded by three blocks of the same size which form the template area. This 4 block set is a square whose sides have length H. Choosing the most relevant neighbourhood to predict the current block is a two step process. First the causal neighbourhood is divided in search windows (two for this 4x4 prediction example) which both have the same height (H) and width (W). From these search areas, blocks are extracted, each block being one pixel away from its neighbours. They are linearized, rows by rows, to get a vector, then normalized and finally added as an atom to a dictionary A. Once this first step is complete, the selection of the closest atom is analogous to the very first step of the method described in [4]. Indeed, the area containing the block to be predicted and the template surrounding it is linearized to get the vector Y. Y can be approximated by one of the atoms multiplied by a weighting factor. However, some components of the vector Y are first unknown: they are those related to pixels belonging to the block to be predicted. So, rows corresponding to the unknown pixels are deleted in both Y and A in order to respectively get Y_c and A_c. Then the algorithm looks for the atom a_j in A_c which has the highest correlation with Y_c such that:

$$j = arg\ max_i\ (a_i^T Y_c)^2 / a_i^T a_i$$

the computed weighting factor w_j is:

$$w_j = (a_j^T Y_c) / a_j^T a_j$$

Then, a_j with all its components is multiplied by w_j in order to get Y_p, the prediction of the current block. This type of prediction relies on the assumption that predicting the template area leads to a good prediction of the current block.

It is possible to predict a block with a linear combination of a set of atoms which have the highest correlation with Y_c. In that case, the prediction is the sum of these atoms each of them being multiplied by its own weighting factor. Then, this sum is divided by the number of atoms involved in the linear combination.

This approach has been implemented in KTA 2.7 to predict 4x4, 8x8 and 16x16 blocks. It competes with all other existing intra mode predictions in H264/AVC.

2.2 Shape of the template area

Several shapes for the template area exist (cf. section 4). The algorithm presented here takes advantage of a template area which has a L shape surrounding the block to be predicted and which is four pixels large whatever the size of the block as depicted in Figure 2.

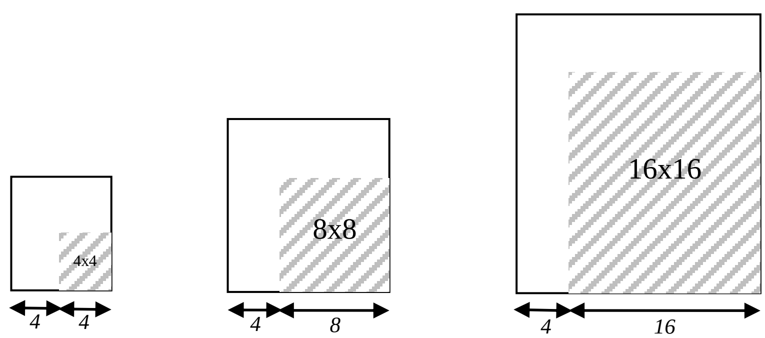

Figure 2 : Size of template area according to block sizes

2.3 Signalling the new intra prediction mode

The syntax of the H264/AVC bit stream was modified in order to specify when this new intra prediction mode is used. Three cases are taken into account: 4x4, 8x8 and 16x16 intra prediction modes.

4x4 and 8x8 intra prediction modes have been modified in a similar way: they both take advantage of the Most Probable Mode (MPM) flag (for instance "prev_intra4x4_pred_mode_flag" in Figure 3). Indeed, if this flag is set, a new second flag, "wtp_flag", must be checked. If this second flag is set, that means that the new intra prediction mode is used, otherwise that means that the most probable mode is used.

mb_pred(mb_type) {		**Descriptor**
if(MbPartPredMode(mb_type, 0) = = Intra_4x4 | | MbPartPredMode(mb_type, 0) = = Intra_8x8 | |		

MbPartPredMode(mb_type, 0) = = Intra_16x16) {		
if(MbPartPredMode(mb_type, 0) = = Intra_4x4)		
for (luma4x4BlkIdx=0; luma4x4BlkIdx<16; luma4x4BlkIdx++) {		
prev_intra4x4_pred_mode_flag [luma4x4BlkIdx**]**		u(1) | ae(v)
if(!prev_intra4x4_pred_mode_flag[luma4x4BlkIdx])		
rem_intra4x4_pred_mode [luma4x4BlkIdx**]**		u(3) | ae(v)
else /* prev_intra4x4_pred_mode_flag**[** luma4x4BlkIdx **]** */		
wtm_flag		u(1) | ae(v)
}		
if(MbPartPredMode(mb_type, 0) = = Intra_8x8)		
for (luma8x8BlkIdx=0; luma8x8BlkIdx<4; luma8x8BlkIdx++) {		
prev_intra8x8_pred_mode_flag [luma8x8BlkIdx**]**		u(1) | ae(v)
if(!prev_intra8x8_pred_mode_flag**[** luma8x8BlkIdx **]**)		
rem_intra8x8_pred_mode [luma8x8BlkIdx**]**		u(3) | ae(v)
else /* prev_intra8x8_pred_mode_flag**[** luma8x8BlkIdx **]** */		
wtm_flag		u(1) | ae(v)
}		
if(MbPartPredMode(mb_type, 0) = = Intra_16x16_Vertical) {		u(1) | ae(v)
wtm_flag		
}		
if (ChromaArrayType = = 1 | | ChromaArrayType = = 2)		
intra_chroma_pred_mode		ue(v) | ae(v)
}		

Figure 3 : Update of 4x4, 8x8 and 16x16 intra prediction syntax

To signal that a 16x16 block was predicted by this new intra mode prediction, the mode 0 (Intra_16x16_Vertical) was extended with a flag. If this flag is set that means that the new intra prediction mode is used, otherwise the classic Intra_16x16_Vertical prediction is used.

2.4 *First application to a test video*

This section presents results of the algorithm described above for a CIF sized test video, “foreman”. All pictures are encoded as Intra and with four different QP values (22, 26, 30, 34) in order to get Bjontegaard metrics[1][10]. They are then compared with those obtained with JM16.2 and KTA2.7. KTA specific settings have been changed and are compliant with those specified in [9]. In particular, Mode Dependent Directional Transform and Adaptative Loop Filter are turned off.

[1] Avsnr release 4 was used.

All 4x4, 8x8 and 16x16 intra prediction H264/AVC modes compete with new intra prediction modes.

The following tables gather first metrics related to the very first picture and then to the whole video. For this test, the number of atoms involved in the linear combination is set to three.

Codec	KTA2.7 + new intra prediction	JM16.2	KTA2.7
PSNR(dB)	+0.36		+0.01
Rate (%)	-6.38		-0.11
Complexity(s)	5.81	6.86	0.98
Relative Complexity	5.9	7	1

Figure 4: First comparisons with JM16.2 and KTA2.7 for the first picture

Codec	KTA2.7 + new intra prediction	JM16.2	KTA2.7
PSNR(dB)	+0.23		+0.03
Rate (%)	-3.56		-0.46
Complexity(s)	1684	2216	307
Relative Complexity	5.48	7.21	1

Figure 5: First comparisons with JM16.2 and KTA2.7 for the whole video

Statistics about selection ratios for this new prediction mode have been computed for the very first picture of this video when QP equals 22 and 34. The following tables gather these figures:

Predicted blocks With the new method	QP=22	QP=34
4x4(%)	11.8	7.4
8x8(%)	10.6	11.6
16x16(%)	9.4	25.8

Figure 6 : Percentage of blocks predicted with the new method

The following picture outlines blocks predicted by the new method in orange:

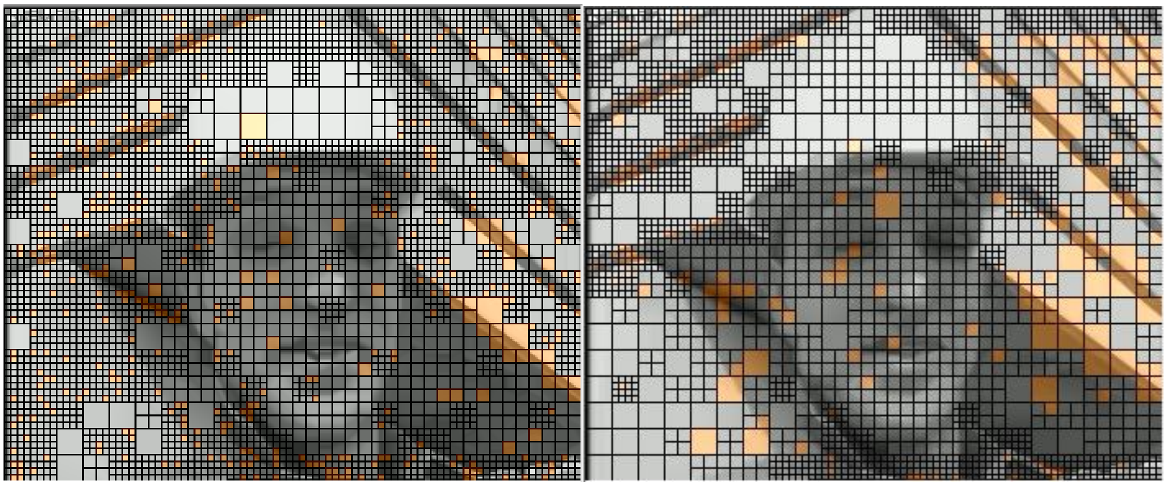

Figure 7 : Example of 4x4, 8x8 and 16x16 blocks predicted when QP=22 (left) and QP=34(right)

3 Experimental conditions and results

This section presents results when videos listed in [9] are encoded with KTA2.7 in which the new intra prediction has been integrated. Results are compared with those obtained with the original KTA2.7. However, in both codecs, MDDT and Adaptative Loop Filer are turned off and encoders have been modified to be compliant with section 4.1 of [9].

QP values are QPI = 22, 26, 30, 34, 38. BD-Rate and BD-PSNR are computed with the lower 4 QP values and the upper 4 QP values.

Results, when all frames to be encoded are Intra, are listed hereafter.

3.1 *Class D, all Intra*

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
BasketballPass	+0.19	-3.05	+0.15	-2.80
BQSquare	+0.05	-0.54	+0.04	-0.47
BlowingBubbles	+0.05	-0.68	+0.04	-0.69
RaceHorses	+0.06	-0.74	+0.05	-0.83
Average	+0.09	-1.25	+0.07	-1.20

3.2 *Class C, all Intra*

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
BasketballDrill	+0.24	-4.71	+0.20	-4.43
BQMall	+0.25	-4.00	+0.26	-4.40
PartyScene	+0.04	-0.48	+0.03	-0.49
RaceHorses	+0.07	-0.88	+0.06	-1.06

Average	+0.15	-2.52	+0.14	-2.60

3.3 Class E, all Intra

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
Vidyo1	+0.25	-4.61	+0.22	-3.82
Vidyo3	+0.20	-3.42	+0.23	-3.84
Vidyo4	+0.11	-2.11	+0.11	-2.18
Average	+0.19	-3.38	+0.19	-3.28

3.4 Class B2, all Intra

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
Cactus	+0.14	-3.35	+0.18	-4.31
BasketballDrive	+0.07	-2.46	+0.09	-3.04
BQTerrace	+0.27	-4.25	+0.28	-5.31
Average	+0.16	-3.35	+0,18	-4.22

3.5 Class B1, all Intra

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
Kimono	+0.03	-0.96	+0.04	-1.04
ParkScene	+0.01	-0.29	+0.01	-0.26
Average	+0.02	-0.63	+0.03	-0.65

3.6 Class A, all Intra

	Lower QP		Upper QP	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
Traffic	+0.07	-1.18	+0.07	-1.25
People on Street	+0.09	-1.46	+0.10	-1.79
Average	+0.08	-1.32	+0.09	-1.52

3.7 First evaluation of complexity

The table below presents relative complexity between KTA2.7 with the new intra prediction and the original KTA2.7 class by class. Note that the new intra prediction uses an external library which is not dedicated to linear algebra and then not optimized. Changing this library should improve the complexity.

Class	KTA2.7 + new intra prediction	KTA2.7
E	22	1
D	12	1
C	14	1
B1+B2	18	1
A	19	1

Figure 8 : Relative complexity between KTA2.7 + new intra prediction and KTA2.7

4 Ongoing and future works

This document has presented a new intra prediction method based on a weighted template matching. It has led to a prediction based on a linear combination of blocks. Note that the signalling could be improved leading to extra gain rate.

Three ongoing works might improve the prediction. First, the number of atoms used in the linear combination has been set to "3" and each of them is determined by one iteration of the matching pursuit algorithm. The performance can be improved by optimizing this number, provided appropriate signalling is implemented. Secondly, the algorithm described in section 2.1 computes a weighting factor thanks to the very first step of a matching pursuit. Going beyond this first iteration, and then selecting the best atom, results in a better prediction. Finally, a third improvement occurs when several shapes of the template area compete with each of them. However, all of these three improvements require an efficient signalling.

Predictions obtained by this algorithm lead to specific residuals which might be efficiently encoded by dedicated transforms, such as Mode-Dependent Directional Transform does for H264/AVC intra prediction residuals [6]. Transforms will be computed thanks to learning techniques such as [7].

All future works will be associated with the integration of this algorithm in the Test Model under Consideration (TMuC) [8].

5 Reference

[1] ISO/IEC JTC1 and ITU-T, "H.264/AVC – Advanced Video Coding for Generic Audiovisual Services", ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 part 10), version 3, 2005.

[2] T.K. Tan, C.S. Boon and Y. Suzuki, "Intra prediction by template matching" ICIP, 2006.

[3] T.K. Tan, C.S. Boon and Y. Suzuki, "Intra prediction by averaged template matching predictors", CCNC, 2007.

[4] A. Martin , J-J Fuchs , C. Guillemot and D. Thoreau, "Atomic decomposition dedicated to AVC and spatial SVC prediction", ICIP, 2008.

[5] http://iphome.hhi.de/suehring/tml/download/KTA/jm11.0kta2.7.zip.

[6] Y. Ye and M. Karczewicz, “Improved H.264 intra coding based on bi-directional intra prediction, directional transform, and adaptive coefficient scanning,” ICIP, 2008.

[7] O. Sezer, O. Harmanci, O. Guleruyz, “Sparse, orthonormal transforms for image compression”, ICIP, 2008.

[8] JCT-VC, “Test Model under Consideration (TMuC)”, 2010.

[9] JCT-VC, “Ad hoc Group on Intra Prediction Overview”, update r1, 2010.

[10] G. Bjontegaard, “Calculation of average PSNR differences between RD-curves”, ITU-T VCEG, 2001.

6 Patent rights declaration(s)

INRIA may have IPR relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).