28th Picture Coding Symposium, PCS2010, December 8-10, 2010, Nagoya, Japan

TOWARDS EFFICIENT INTRA PREDICTION BASED ON IMAGE INPAINTING METHODS

Dimitar Doshkov[1], Patrick Ndjiki-Nya[1], Haricharan Lakshman[1], Martin Köppel[1], and Thomas Wiegand[1,2]

[1]Image Processing Department
Fraunhofer Institute for Telecommunications
Heinrich Hertz Institute, Berlin, Germany

[2]Image Communication Chair
Department of Telecommunication Systems
Technical University of Berlin, Germany

{dimitar.doshkov/patrick.ndjiki-nya/haricharan.lakshman/martin.koeppel/thomas.wiegand}@hhi.fraunhofer.de

ABSTRACT

In this paper, novel intra prediction methods based on image inpainting approaches are proposed. The H.264/AVC intra prediction modes are not well suited for processing complex textures at low bit rates. Our algorithm utilizes an efficient combination of partial differential equations (PDEs) and patch-based texture synthesis in addition to the standard directional predictors. Bit rate savings up to 3.5% compared to that of the H.264/AVC standard are shown.

Index Terms— Intra prediction, Texture synthesis, PDEs, Template matching, Inpainting

1. INTRODUCTION

H.264/AVC [1] is a hybrid video coding standard with block-based architecture. It utilizes two important prediction modes: Intra and Inter. The encoder chooses the best mode for the current block in terms of rate-distortion (RD) performance and transmits side information in the bitstream. The decoder builds the same prediction signal using the received information and reconstructs the block. Intra prediction is a key tool in video compression to exploit the spatial correlation within a picture. The prediction of the current block is formed by using the neighboring already reconstructed samples (A-M, cf. Fig. 1a). This is performed for different block sizes: 16×16, 8×8 and 4×4, etc. H.264/AVC supports eight directional modes (cf. Fig. 1b) plus the DC mode for 4×4 and 8×8 blocks and four modes (vertical, horizontal, DC and plane prediction) for 16×16 blocks.

In Computer Vision, the seamless regeneration of missing information in an image is usually referred to as inpainting. Texture synthesis (TS) is an appropriate inpainting approach for generating large regions with complex textures, for instance, using patch based template matching [2],[3]. Inpainting based on solving partial differential equations (PDEs) [4] is more suitable for restoration of small rather homogenous regions. Template matching has been already applied to intra frame coding in [5],[6] and [7]. Tan et al. [5] proposed a basic template matching approach to obtain a prediction of the current block using texture synthesis. In [6], the authors improved the method in [5] by utilizing predictor averaging. Motivated by the confidence map based technique in [2], Guo et al. [7] proposed the advantage of the priority-based template matching intra prediction over the non-priority-based algorithm in [5]. Image inpainting methods that rely on the well-known Laplace PDE have also provided effective methods for image [8] and video [9] compression. The work in [8], [9] explicitly detect and transmit edges and use Laplace PDE only for smooth regions. However, accurate edge detection is a difficult and unreliable operation in the context of video coding.

In this paper, we avoid explicit edge detection by designing directional predictors using the concept of Laplace PDE. The advantage compared to the H.264/AVC directional modes is that the proposed approach also considers homogeneity in the spatial neighborhood while predicting samples in predetermined directions. Furthermore, different methods of inpainting (based on PDEs and template matching) are combined to obtain new intra predictors suitable for all block sizes.

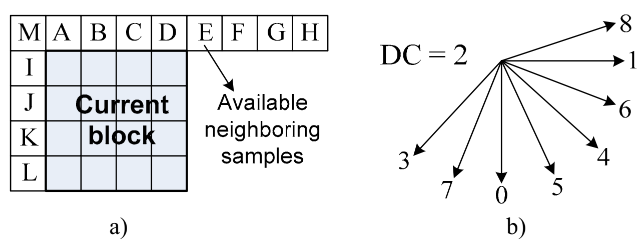


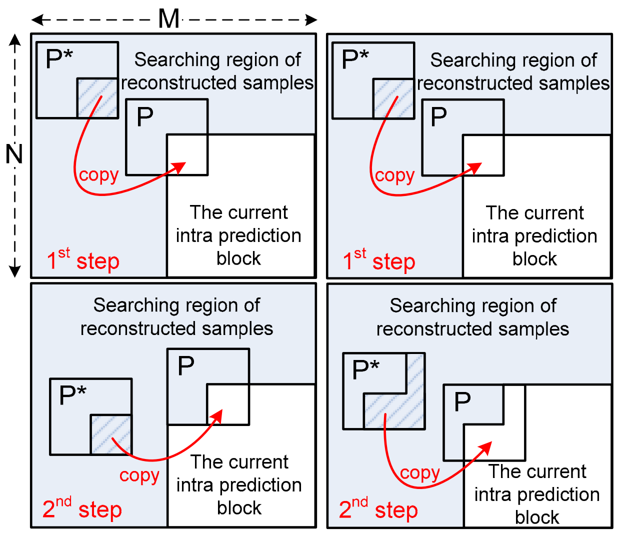
Fig. 1. *a) Intra prediction setting for 4×4 block size in H.264/AVC, b) corresponding directional modes.*

2. INTRA PREDICTION BY TEMPLATE MATCHING

In this section, we briefly describe both non-priority [5] and priority [7] based template matching for intra prediction and then present our proposal.

2.1. State of the Art

In the works by Tan et al. [5] and Guo et al. [7], texture synthesis methods using template matching are proposed to predict the missing information in intra blocks. Tan et al. [5] first divide 4×4 or 8×8 blocks into 2×2 sub-blocks. Then, using the neighboring samples above and to the left of the sub-block (samples of region P in Fig. 2), the best candidate is searched in the closed area M×N of the already reconstructed samples. A Sum of Absolute Differences (SAD) between the templates P and P* is used as a matching criterion. Note that, this texture synthesis approach assigns the output samples in a raster scan order (cf. Fig. 2a, 2nd step).



a) b)

Fig. 2. *Principle of a) non-priority-based and b) priority-based template matching.*

In [7], a priority based template matching for predicting the current block from the region of reconstructed samples is presented. This method estimates the current block in the following manner (please refer to [2] and [7] for a detailed explanation of the original algorithm): 1) The priority values of the border samples of the current intra block are calculated; 2) The border sample with the highest priority is selected as the center of the next template; 3) Finally, a search procedure is conducted to find the best template candidate in the region of reconstructed samples. The measure of the best match is the same as for the non-priority approach.

Fig. 2 shows the difference between the two template matching methods. It can be seen, that the fill-in order (cf. Fig. 2 a,b; 2nd step) is different for each approach. The main advantage of the priority-based template matching lies in its ability to deal with regions with mixed structure and texture [7].

2.2. Improved Averaged Template Matching

The proposed method is motivated by the work by Tan et al. [6] in which it was shown that averaging N candidate templates (patches) results in a better sample prediction than using only one template during the fill-in process. Our algorithm (TS_{NEW}) represents a two candidate templates method where a combination of both priority (TS_P) and non-priority (TS_{NP}) template matching is used. Firstly, the current block to-be-coded (cf. Fig. 2) is predicted separately with the two methods described above. This step produces two different, but perceptually similar results. Note that, both synthesis approaches apply the same conditions regarding search area ($M \times N$) and patch size to establish the block to-be-coded. Furthermore, we apply the Sum of Squared Errors (SSE) instead of SAD as a matching criterion. This is done in order to have a single distortion metric for both template matching and RD optimization. Secondly, the average of both predictors is built, i.e. the final prediction result of the current block is the average of priority and non-priority template matching. In comparison to our implementation, the strategy of Tan et al. [6] applies only the non-priority method and the averaging is done during the texture synthesis, i.e. online at patch level, not at block level.

3. INTRA PREDICTION BY PDE-BASED INPAINTING

In this section, we consider the Laplace equation as an appropriate approach to perform an intra prediction. This inpainting technique represents the point of origin of our further new prediction modes, which we provide to enhance coding efficiency of H.264/AVC.

3.1. Overview of Standard Laplace PDE-based Inpainting

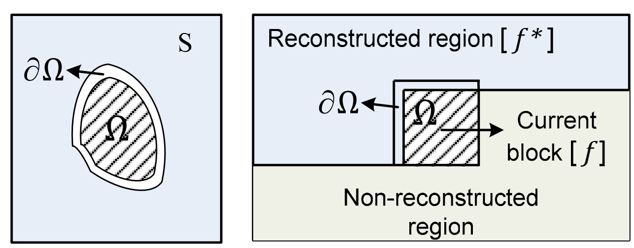
In order to clarify the inpainting problem, we define f^* as a known scalar function over the domain S ($S \subset \mathbb{R}^2$). As indicated in Fig.3a, $\partial\Omega$ represents the boundary of the unknown area $\Omega \subset S$. f is an unknown scalar function defined over Ω. The aim is to find f using only the information available in $\partial\Omega$. This boundary value problem can be expressed as:

$$\Delta f = 0 \tag{1}$$

with the Dirichlet boundary condition

$$f|_{\partial\Omega} = f^*|_{\partial\Omega}\,, \tag{2}$$

where, Δ represents the Laplacian operator. In this way, information on the boundary $\partial\Omega$ is diffused into Ω, such that the final result is smooth.



a) b)

Fig. 3. *a) General case of the Laplace inpainting approach. b) Notations for intra prediction with Laplace inpainting.*

3.2. Novel PDE-based Directional Prediction

The notations of Laplace inpainting, in the context of intra prediction, are illustrated in Fig. 3b. It can be seen that minimum two of the four sides of the block to-be-coded are always adjacent to the non-reconstructed area. In this case, the current block is represented by f, the already reconstructed region is denoted as $f^*$and the boundary $\partial\Omega$ comprises the spatial neighbors above and to the left of the current block. The boundary conditions of Laplace equation for the region Ω are incomplete because of the presence of non-reconstructed area. Therefore we adapt the Laplace equation to use only the available samples in the considered region. We denote the numerical approximation of the 2D-Laplace operator as:

$$\Delta \cong \begin{bmatrix} 0 & -C_N & 0 \\ -C_W & C_C & -C_E \\ 0 & -C_S & 0 \end{bmatrix} \text{ with } \begin{cases} C_{(i)_{i=\{N,E,S,W\}}} = 1 \\ C_c = 4 \end{cases}. \tag{3}$$

Then, the Laplace equation (1) can be estimated as

$$\sum_{i=\{N,E,S,W\}} w_{(i)}(f_C - f_{(i)}) = 0 \quad \text{with} \quad f_C \in \Omega, \tag{4}$$

where, w represents an indicator function that gives 1 for available samples and 0 for unavailable, i.e.:

$$w_{(i)}{}_{i=\{N,E,S,W\}} = \begin{cases} 1, & \text{if } i \in \Omega \cup \partial\Omega \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The samples of region Ω can be obtained by solving the system of linear equations according to equation (4).

It can be seen that the coefficient values of the matrix Δ defined in equation (3) results in an estimation of region Ω such that each sample is a local average of its neighboring samples. Therefore it is suitable for estimating smooth regions. We propose an innovative approach to control the prediction direction at the same time maintaining some amount of spatial homogeneity by modifying the 2D-Laplace operator. Here, we use the vertical prediction as an example to explain the new scheme. The derivation of the other predictors is analogous. Using the 3×3 matrix notation of equation (3), we can define vertical intra prediction as $C_N = C_C = 1$ and $C_E = C_S = C_W = 0$. It can be also presented as the gradients of f in the y-direction $(grad f_y)$. Now, the problem of estimating f with respect to the vertical trend, equation (1) may be converted to the following boundary value problem:

$$\Delta f - grad f_y = 0. \tag{6}$$

Equation (6) forms an elegant combination of two cost functions, Laplace inpainting and vertical intra prediction. Using the elucidations in (3) and (4), we can write the discrete form of (6) as:

$$\sum_{i=\{N,E,S,W\}} w_{(i)}(f_C - f_{(i)}) - w_N(f_N - f_C) = 0 \quad \text{with} \quad f_C \in \Omega. \tag{7}$$

After some simplifications, it can be seen that equation (7) takes the form of equation (3) when $C_N = 2$, $C_E = C_S = C_W = 1$ and $C_C = 5$. Table 1 summarizes all new intra modes with the corresponding coefficients for the 3×3 mask. Fig. 4 shows examples of the visual difference between the proposed intra prediction PDE-based methods and the corresponding modes of H.264/AVC.

Table 1. *Coefficients of the 3×3 mask for PDE-based directional prediction.*

Direction		Mask				
		C_N	C_E	C_S	C_W	C_C
Inpaint. Vertical	(I_V)	2	1	1	1	5
Inpaint. Horizontal	(I_H)	1	1	1	2	5
Laplace Inpaint.	(I_L)	1	1	1	1	4
Inp. Diag. Down-Left	(I_DDL)	2	2	1	1	6
Inp. Vertical-Right	(I_VR)	3	2	1	3	9
Inp. Horizontal-Down	(I_HD)	3	1	2	3	9
Inp. Horizontal-Up	(I_HU)	1	1	2	2	6

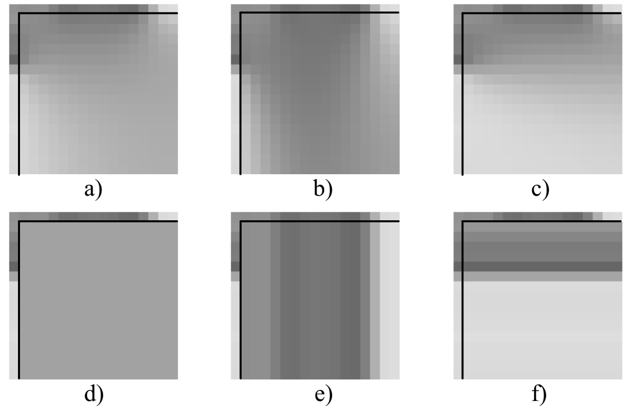


Fig. 4. *Comparison between PDE-based directional prediction and H.264/AVC prediction modes by means of 16×16 block size. All blocks are shown with the original spatial neighbors (one sample width, top row and outmost left column). Results of the PDE-based prediction with a) Laplace (I_L), b) vertical (I_V), c) horizontal (I_H) inpainting. Results of H.264/AVC d) DC, e) vertical and f) horizontal prediction.*

4. EXPERIMENTAL CONDITIONS AND RESULTS

We integrated the new intra predictors into the H.264/AVC reference software, version JM16.2 [11]. We use six video sequences with a frame rate of 30Hz (cf. Table 3) in these evaluations. The experiments are conducted with high profile encoder settings, including 8×8 transforms, CABAC entropy coding [1] and RD optimized mode decision. Furthermore, all test video sequences are intra coded at QP levels 28, 32, 36 and 40. The average luma PSNR improvement as well as the bit rate savings are calculated using the Bjøntegaard metric [10]. It has been observed that it is adequate to set the search area M×N (cf. Fig. 2) to M=N=32 and the patch size (PS) for all texture synthesis approaches to PS=7 for intra 16×16; M=N={16, 24} (depending on the available spatial neighbor blocks), PS=5 for intra 8×8 and M=N={12, 16}, PS=5 for intra 4×4.

Table 2. *Specification of the integration of the new intra modes.*

Mode	Mode Integration	Abbreviation
16×16	H.264- original modes [1]	16_H.264
	16_H.264 **&** TS_{NEW} instead of Plane	16_TS_{NEW}
	16_H.264 **&** I_L instead of DC-mode **&** I_H instead of Horizontal mode **&** I_V instead of Vertical mode	16_I
	16_TS_{NEW} **&** 16_I	16_Opt
4×4 & 8×8	H.264- original modes [1]	4_8_H.264
	4_8_H.264 **PLUS** TS_P	4_8_TS_P
	4_8_H.264 **PLUS** TS_{NP}	4_8_TS_{NP}
	4_8_H.264 **PLUS** TS_{NEW}	4_8_TS_{NEW}
	4_8_H.264 **PLUS** all Inp. modes in Table 1	4_8_I
4×4 & 8×8 & 16×16	16_H.264 **&** 4_8_TS_P	4_8_16_TS_P
	16_H.264 **&** 4_8_TS_{NP}	4_8_16_TS_{NP}
	16_H.264 **&** 4_8_TS_{NEW}	4_8_16_TS_{NEW}
	16_I **&** 4_8_I	4_8_16_I
	16_Opt **&** 4_8_TS_{NEW} **&** 4_8_I	4_8_16_Opt

Table 3. *Bit rate differences of new and state-of-the-art intra modes in comparison to H.264/AVC.*

Video	Resolution	**Comparison 1 (only 16×16 is enabled)**			**Comparison 2 (all modes are enabled)**				
		16_TS_{NEW}	16_I	16_Opt	$4\_8\_16\_TS_{P}$	$4\_8\_16\_TS_{NP}$	$4\_8\_16\_TS_{NEW}$	4_8_16_I	4_8_16_Opt
BQSquare	416×240	-0.04 %	-0.57 %	-0.62 %	0.93 %	0.92 %	0.69 %	0.41 %	-0.67 %
RHorses		-0.04 %	-3.48 %	-3.98 %	1.30 %	1.31 %	1.04 %	-0.38 %	-1.35 %
Foreman	352×288	-13.46 %	-0.41 %	-13.38 %	0.61 %	0.32 %	-0.91 %	2.27 %	-3.49 %
Mobile		-0.08 %	-0.32 %	-0.55 %	0.14 %	0.22 %	-0.03 %	-0.34 %	-1.40 %
Paris		-0.56 %	5.16 %	1.23 %	-0.22 %	-0.24 %	-0.57 %	0.10 %	-1.91 %
Tempete		0.04 %	-1.40 %	-2.30 %	1.18 %	1.27 %	1.06 %	-1.07 %	-1.35 %
Average		-2.36 %	-0.17 %	-3.27 %	0.66 %	0.63 %	0.21 %	0.16 %	-1.70 %

We realized the new intra modes for 4×4, 8×8 and 16×16 blocks. It was found that the new predictors worked well as additional modes for 4×4 and 8×8 blocks and replacement of the 4 existing modes for 16×16. Note that the signaling of the additional modes is done with 4 bits instead of 3 bits in H.264/AVC.

We conducted two main comparisons. The test conditions are summarized in Table 2, while Table 3 lists representative results of the comparisons. All provided average bit rate differences are always estimated in comparison to H.264/AVC, with 35 frames for "Comparison 1" and with 15 frames for "Comparison 2" (cf. Table 3).

From "Comparison 1" in Table 3, it can be seen, that the results depend on the video characteristics. In general, for videos such as Foreman, which contain long and clean edges, TS_{NEW} has larger gains in comparison to the PDE-based approaches (16_I). In some cases PDE-based methods give better results (eg. RHorses). When allowing both methods for 16×16 blocks (cf. Table 3, 16_Opt, "Comparison 1"), an enhanced average of 3.27% bit rate saving which translates into 0.173 dB of PSNR gain is achieved.

We also implemented the new methods for 4×4 and 8×8 blocks as additional modes. "Comparison 2" gives the objective bit rate results. The test conditions $4\_8\_16\_TS_P$ and $4\_8\_16\_TS_{NP}$ correspond to the work of [7] and [5] respectively. It should be noted that we use a different syntax (parameter, signaling) and software version (newest version used in this work) than [7] and [5], which may explain the fact that the estimated bit rate differences for $4\_8\_16\_TS_P$ and $4\_8\_16\_TS_{NP}$ are not the same as those presented in [7] and [5]. Columns $4\_8\_16\_TS_{NEW}$ and 4_8_16_I (cf. "Comparison 2" in Table 3) report the bit rate differences of the separate integration of the proposed methods. Table 3 shows that enabling all new intra modes to 4_8_16_Opt yields more than 3.5% bit rate saving for Foreman and shows an average 1.7% bit rate saving which translates into 0.108 dB of PSNR gain overall.

A downside of the proposed new intra prediction modes is the high computational complexity at both the encoder and the decoder. The main complexities are related to the time-expensive template based texture synthesis approaches and the PDEs solving step of the inpainting methods. However, the aim of this work is to demonstrate the potential of the presented intra modes in terms of coding efficiency.

5. CONCLUSION

This paper presented a new H.264/AVC intra coding method that combines two inpainting approaches (PDE-based and template matching). The new intra mode shows potential to improve the intra prediction in terms of coding efficiency. In future work, we will further analyze the tradeoff between the mode complexity and the RD performance. Further tasks will involve research towards new context-adaptive inpainting based intra prediction methods.

6. REFERENCES

[1] "Advanced Video Coding for Generic Audiovisual Services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC)," Standard Version 7, ITU-T and ISO/IEC JTC 1, Apr.2007.

[2] A. Criminisi, P. Perez, and K. Toyama, "Region Filling and Object Removal by Exemplar-based Image Inpainting," *IEEE Trans. on Image Processing*, vol. 13, no. 9, pp. 1200-1212, Sep. 2004.

[3] P. Ndjiki-Nya, M. Köppel, D. Doshkov, Thomas Wiegand, "Automatic Structure-Aware Inpainting for Complex Image Content," *Pros. ISVC*, Las Vegas, NV, USA, pp. 1144-1156, Dec. 2008.

[4] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting," *Proc. ACM SIGGRAPH 2000*, New Orleans, pp. 417-424, Jul. 2000.

[5] T. K. Tan, C. S. Boon, and Y. Suzuki, "Intra Prediction by Template Matching," *Proc. ICIP*, Atlanta, GA, USA, pp. 1693-1696, Oct. 2006.

[6] T. K. Tan, C. S. Boon, and Y. Suzuki, "Intra Prediction by Averaged Template Matching Predictors," *Proc. CCNC*, Las Vegas, Nevada, USA, pp. 405-409, Jan. 2007.

[7] Y. Guo, Y.-K. Wang, and H. Li, "Priority-Based Template Matching Intra Prediction," *Proc. ICME*, Hannover, Germany, pp. 1117-1120, June 2008.

[8] V. Bastani, M. S. Helfroush, and K. Kasiri, "Image Compression based on Spatial Redundancy Removal and Image Inpainting," *Zhejiang University Press, co-published with Springer*, pp. 91-100, Jan. 2010.

[9] D. Liu, X. Sun, F. Wu, and Y.-Q. Zhang, "Edge-oriented Uniform Intra Prediction," *IEEE TIP*, vol. 17, no. 10, pp. 1827-1836, Oct. 2008.

[10] S. Pateux, and L. Jung, "An Excel Add-in for Computing Bjontegaard Metric and Its Evolution," *VCEG-AE07*, Jan. 2007.

[11] H.264/AVC reference software (the Joint Model-JM), version 16.2, available at http://iphome.hhi.de/suehring/tml/download/old_jm/jm16.2.zip