ABSTRACT

The paper compares a sparse approximation based spatial texture prediction method with the template matching based prediction. Template matching algorithms have been widely considered for image prediction. These approaches rely on the assumption that the predicted texture contains a similar textural structure with the template in the sense of a simple distance metric between template and candidate. However, in real images, there are more complex textured areas where template matching fails. The basic idea instead is to consider sparse approximation algorithms. The proposed sparse spatial prediction is assessed against the prediction method based on template matching with a static and optimized dynamic templates. The spatial prediction method is then assessed in a coding scheme where the prediction residue is encoded with a coding approach similar to JPEG. Experimental observations show that the proposed method outperforms the conventional template matching based prediction.

Index Terms— Texture prediction, sparse approximation, matching pursuits, template matching, dynamic template

1. INTRODUCTION

Closed-loop intra prediction plays an important role in minimizing the encoded information of an image or an intra frame in a video sequence. E.g., in H.264/AVC, there are two intra prediction types called Intra-16x16 and Intra-4x4 respectively [1]. The Intra-16x16 type supports four intra prediction modes while the Intra-4x4 type supports nine modes. Each 4x4 block is predicted from prior encoded samples from spatially neighboring blocks. In addition to the so-called “DC” mode which consists in predicting the entire 4x4 block from the mean of neighboring pixels, eight directional prediction modes are specified. The prediction is done by simply “propagating” the pixel values along the specified direction. This approach is suitable in presence of contours, when the directional mode chosen corresponds to the orientation of the contour. However, it fails in more complex textured areas.

An alternative spatial prediction algorithm based on template matching has been described in [2]. In this method, the block to be predicted of size 4x4 is further divided into four 2x2 sub-blocks. Template matching based prediction is conducted for each sub-block accordingly. The best candidate sub-block of size 2x2 is determined by minimizing the sum of absolute distance (SAD) between template and candidate neighborhood. The four best match candidate sub-blocks constitute the prediction of the block to be predicted. This approach has later been improved in [3] by averaging the multiple template matching predictors, including larger and directional templates, as a result of more than 15% coding efficiency in H.264/AVC. Any extensions and variations of this method are straightforward. In the experiments reported in this paper, 8x8 block size has been used without further dividing the block into sub-blocks.

Here, a spatial prediction method based on sparse signal approximation (such as matching pursuits (MP) [4], orthogonal matching pursuits (OMP) [5], etc.) is considered and assessed competitively to the template matching based spatial prediction technique. The principle of the approach, as initially proposed in [6], is to first search for the linear combination of basis functions which best approximates known sample values in a causal neighborhood (template), and keep the same linear combination of basis functions to approximate the unknown sample values in the block to be predicted. Since a good representation of the template does not necessarily lead to a good approximation of the block to be predicted, the iteration number, which minimizes a chosen criterion, needs to be transmitted to the decoder. The considered criteria are the mean square error (MSE) of the predicted signal and a rate-distortion (RD) cost function when the prediction is used in a coding scheme.

Note that, this approach can be seen as an extension of the template matching based prediction (which keeps one basis function with a weighting coefficient equal to 1). In order to have a fair comparison with template matching, the sparse prediction algorithm is iterated only once. In the experiments reported here, the OMP algorithm has been used by considering a locally adaptive dictionary as defined in [6]. In addition, both a static and MSE/RD optimized dynamic templates are used. The best approximation support (or template) among a set of seven pre-defined templates is selected according to the corresponding criterion, that is minimizing either the residual MSE or the RD cost function on the predicted block.
The proposed spatial prediction method has been assessed in a coding scheme in which the residue blocks are encoded with an algorithm similar to JPEG. The approximation support type (if dynamic templates are used) is Huffman encoded. The prediction and coding PSNR/bit-rate performance curves show a gain up to 3 dB when compared with the conventional template matching based prediction.

2. SPATIAL PREDICTION

Let \( S \) denote a region in the image containing a block \( B \) of \( n \times n \) pixels and its causal neighborhood \( C \) used as approximation support (template) as shown in Fig. 1. The region \( S \) contains 4 blocks, hence of size \( N = 2n \times 2n \) pixels, for running the prediction algorithm. In the region \( S \), there are known values (the template \( C \)) and unknowns (the values of the pixels of the block \( B \) to be predicted). The principle of the prediction approach is to first search for the best approximation for the known values in \( C \), and keep the same procedure to approximate the unknown pixel values in \( B \).

The \( N \) sample values of the area \( S \) are stacked in a vector \( b \). Let \( A \) be the corresponding dictionary for the prediction algorithm represented by a matrix of dimension \( N \times M \), where \( M \geq N \). The dictionary \( A \) is constructed by stacking the luminance values of all patches in a given causal search window \( W \) in the reconstructed image region as shown in Fig. 1. The use of a causal window guarantees that the decoder can construct the same dictionary.

2.1. Template matching (TM) based spatial prediction

Given \( A \in \mathbb{R}^{N \times M} \) and \( b \in \mathbb{R}^N \), the template matching algorithm searches the best match between template and candidate sample values. The vector \( b \) is known as the template (also referred as the filter mask), and the matrix \( A \) is referred as the dictionary where its columns \( a_j \) are the candidates. The candidates correspond to the luminance values of texture patches extracted from the search window \( W \).

The problem of template matching seeks a solution to minimization of a distance \( d \) as

\[
\arg \min_{j \in \{1 \ldots M\}} \{ d_j : d_j = \text{DIST}(b, a_j) \}.
\]

Here, the operator \text{DIST} denotes a simple distance metric such as sum of squared distance (SSD), SAD, MSE, etc. The best match (minimum distance) candidate is assigned as the predictor of the template \( b \).

2.1.1. Static template prediction

A static template is referred as the commonly used conventional template, i.e., mode 1 in Fig. 2. Let us suppose that the static template (mode 1) is used for prediction. For the first step, that is search for the best approximation of the known pixel values, the matrix \( A \) is modified by masking its rows corresponding to the spatial location of the pixels of the area \( B \) (the unknown pixel values). A compacted matrix \( A_c \) of size \( 3n^2 \times M \) is obtained. The known input image is compacted in the vector \( b_c \) of \( 3n^2 \) values.

Let \( a_{c_j} \) denote the columns of the compacted dictionary \( A_c \). The template matching algorithm proceeds by calculating \( d_j = \text{DIST}(b_c, a_{c_j}) \) for all \( j = 1 \ldots M \) in order to obtain

\[
j_{\text{opt}} = \arg \min_j \{ d_j \}
\]

The extrapolated signal \( \hat{b} \) is simply assigned by the sample values of the candidate \( a_{c_{j_{\text{opt}}}} \) as \( \hat{b} = a_{c_{j_{\text{opt}}}} \).

2.1.2. Optimized dynamic templates

The optimum dynamic template is selected among seven predefined modes as shown in Fig. 2. The optimization is conducted according to two different criteria: 1. minimization of the prediction residue MSE; 2. minimization of the RD cost function \( J = D + \lambda R \), where \( D \) is the reconstructed block MSE (after adding the quantized residue when used in the coding scheme), and \( R \) is the residue coding cost estimated as \( R = \gamma_0 M' \) [7] with \( M' \) being defined as the number of non-zero quantized DCT coefficients and \( \gamma_0 = 6.5 \).
2.2. Sparse approximation based spatial prediction

Given \( A \in \mathbb{R}^{N \times M} \) and \( b \in \mathbb{R}^N \) with \( N << M \) and \( A \) is of full rank, the problem of sparse approximation consists in seeking the solution of

\[
\min \{ ||x||_0 : Ax = b \},
\]

where \( ||x||_0 \) denotes the \( L_0 \) norm of \( x \), i.e., the number of non-zero components in \( x \). \( A \) is known as the dictionary, its columns \( a_j \) are the atoms, they are assumed to be normalized in Euclidean norm. There are many solutions \( x \) to \( Ax = b \) and the problem is to find the sparsest, i.e., the one for which \( x \) has the fewest non-zero components.

In practice, one actually seeks an approximate solution which satisfies:

\[
\min \{ ||x||_0 : \|Ax - b\|_p \leq \rho \},
\]

for some \( \rho \geq 0 \), characterizing an admissible reconstruction error. The norm \( \rho \) is usually 2. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary \( A \). Searching for this sparsest representation is hence computationally intractable. MP algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

In this paper, the OMP algorithm which offers an iterative optimal solution to the above problem is considered. It generates a sequence of \( M \) dimensional vectors \( x_k \) having an increasing number of non-zero components in the following way. At the first iteration \( x_0 = 0 \) and an initial residual vector \( r_0 = b - Ax_0 = b \) is computed. At iteration \( k \), the algorithm identifies the atom \( a_{j_k} \) having the maximum correlation with the approximation error. Let \( A_{jk} \) denote the matrix containing all the atoms selected in the previous iterations. One then projects \( b \) onto the subspace spanned by the columns of \( A_{jk} \), i.e., one solves

\[
\min_x \| A_{jk} x - b \|_2,
\]

and the coefficient vector at the \( k \)th iteration is given as

\[
x_k = (A_k^T A_k)^{-1} A_k^T b = A_k^+ b,
\]

where \( A_k^+ \) is the pseudo-inverse of \( A_k \). Notice that here \( x_k \) is a vector of coefficients. All the coefficients assigned to the selected atoms are recomputed at each step. However, in the experiments reported here, only one iteration has been considered for the sake of comparison with template matching.

### 2.2.1. Application to prediction

The principle of the sparse approximation approach is to first search for a best basis function (atom) which best approximates the known values in \( C \), and keep the same basis function and weighting coefficient to approximate the unknown pixel values in \( B \).

---

**Fig. 3.** Prediction results of Foreman image.

Let the quantity \( x \) denote a vector which will contain the result of the sparse approximation, i.e., the coefficient of the expansion of the vector \( b \) on only one atom. The sparse representation algorithm then proceeds by solving the approximate minimization

\[
x_{opt} = \min_x \| b_k - A_k x \|_2 \quad \text{subject to} \quad ||x||_0 = 1.
\]

To recover the extrapolated signal \( \hat{b} \), the full matrix \( A \) is multiplied by \( x_{opt} \) as \( \hat{b} = Ax_{opt} \).

Here, the columns (atoms) \( a_{c_j} \) of the compacted dictionary \( A_k \) correspond to the normalized luminance values of texture patches extracted from the search window \( W \).

### 3. EXPERIMENTAL RESULTS

The proposed sparse spatial prediction method has been assessed competitively to the template matching based prediction using both a static and optimized dynamic templates. It has then been assessed in a still image coding scheme in which the residue blocks are encoded with an algorithm similar to JPEG.

In order to initialize the prediction, the top 3 rows and left 3 columns of blocks of size \( 8 \times 8 \) are intra coded with JPEG algorithm. When a block has been predicted, the residue is quantized and encoded similar to JPEG. In this coding structure, a uniform quantization matrix of step size equal to 16 is weighted by a quality factor (QF). The QF is increased from 10 to 90 with a step size of 10. The reconstructed image is obtained by adding the quantized residue to the prediction. The
best approximation support (if dynamic template selection is used) is Huffman encoded when transmitted to the decoder. For the sake of the comparison with template matching, the OMP algorithm is iterated only once.

Fig. 3 shows the predicted images of Foreman1 at QF=30 where the dynamic template selection criterion is the minimization of the prediction residue MSE. Fig. 4 demonstrates the corresponding prediction performance curves for Foreman (CIF) and Barbara (512 × 512) images. The quality of the predicted signal, both visually and in terms of PSNR/bitrage, is significantly improved when compared with the conventional template matching based prediction.

Fig. 5 shows the coding performance curves of the proposed sparse image prediction algorithm in comparison to the template matching based prediction using both a static and RD optimized dynamic templates. One can observe that the proposed method with dynamic template selection significantly improves the coding performance with respect to the conventional template matching prediction. A gain of up to 3 dB has been achieved in both prediction and coding.

4. CONCLUSION

In the paper, a sparse spatial texture prediction method has been introduced. The proposed prediction method with dynamic templates offers an effective alternative for minimizing the encoded information of an image or an intra frame in comparison to conventional template matching based prediction. This method can be regarded as a correlation based template matching with dynamic templates where weighting coefficient may not equal to 1. Since the approximation algorithm has been iterated once, the computational complexity can be neglected when compared to template matching.

5. REFERENCES