Sequential Error Concealment for Video/Images by Sparse Linear Prediction

Koloda, Jan; Østergaard, Jan; Jensen, Søren Holdt; Sanchez, Victoria; Peinado, Antonio

Published in:
I E E E Transactions on Multimedia

Publication date:
2013

Document Version
Early version, also known as pre-print

Sequential Error Concealment for Video/Images by Sparse Linear Prediction

Ján Koloda, Jan Østergaard, *Senior Member, IEEE,* Søren H. Jensen, *Senior Member, IEEE,* Victoria Sánchez, *Member, IEEE* and Antonio M. Peinado, *Senior Member, IEEE*

***Abstract*—In this paper we propose a novel sequential error concealment algorithm for video and images based on sparse linear prediction. Block-based coding schemes in packet loss environment are considered. Images are modelled by means of linear prediction and missing macroblocks are sequentially reconstructed using the available groups of pixels. The optimal predictor coefficients are computed by applying a missing data regression imputation procedure with a sparsity constraint. Moreover, an efficient procedure for the computation of these coefficients based on an exponential approximation is also proposed. Both techniques provide high quality reconstructions and outperform the state-of-the-art algorithms both in terms of PSNR and MS-SSIM.**

***Index Terms*—Error concealment, block-coded images/video, convex optimization, missing data imputation, sparse representation**

I. Introduction

BLOCK-based video coding standards, such as MPEG-4 or H.264/AVC, are widely used in multimedia applications. Video signals are split into macroblocks that are coded using inter- or intraframe prediction. Quantization is carried out in the DCT domain and lossless arithmetic compression is applied [1]. This leads to low distortions at moderate bit-rates. However, achieving high quality reception is a challenging task since data streams are usually transmitted over error-prone channels.

For real-time transmission applications, the H.264/AVC standard has introduced several error resilience tools, such as arbitrary slice order (ASO) and flexible macroblock ordering (FMO) [2]. Macroblocks within a frame can be split into several slices. A slice forms the payload of a network abstraction layer unit (NALU), which is a data sequence that can be decoded independently [1]. The loss of a NALU will therefore not affect other macroblocks within the current frame. However, due to temporal interframe prediction, error propagation does occur.

H.264/AVC allows both bit- and packet-oriented delivery. For bit-oriented transmissions, an error burst that surpasses the channel-coding protection may result in loss of synchronization as well as fatal data damage since H.264/AVC utilizes variable length coding (VLC) or Exponential-Golomb coding for lossless compression [3]. Errors would thus propagate throughout the packet, making the current slice unusable. In packet oriented delivery, damaged packets, containing NALUs, are usually detected and discarded by network or transmission layers. Also, there may be packets which are not received at all due to congestion, routing problems, etc. In both cases, we are facing the problem of the loss of, at least, one slice.

Error concealment (EC) techniques form a very challenging field, since QoS is of utmost importance for the users. In many cases, retransmission of lost data is not possible due to real-time constraints or lack of bandwidth. This last case also applies to additional transmission of media-specific forward error correction (FEC) codes which, in addition, may not be standard compliant [4]. In contrast to channel coding techniques, which are carried out at the encoder and are designed to minimize the negative impact of packet losses, EC is applied at the decoder and can significantly improve the quality of the received stream [5]. EC algorithms can be classified into two categories: spatial EC (SEC), which relies on the information provided within the current frame and temporal EC (TEC), which utilizes temporal information such as motion vectors (MV) and previous or already available future frames. Some TEC techniques use both temporal and spatial information for image restoration and they are often referred to as combined or hybrid SEC/TEC algorithms. Both categories, SEC and TEC, exploit the redundancy due to the high spatial and temporal correlation within a video sequence. Temporal correlations tend to be higher than the spatial ones, so TEC techniques usually provide better results. This would be the straightforward choice when concealing a P/B-frame (intercoded). However, utilizing temporal information for the recovery of I-frames (intracoded) is not always possible, since they may be inserted to reset the prediction error when a change of scene occurs. Thus, when all the available temporal information belongs to a different scene or there is no temporal information available, SEC algorithms are necessary. Every I/P-frame in the video sequence usually serves as a prediction template for, at least, one intercoded frame. Thus, high quality concealment is required since any reconstruction error will be propagated until the next I-frame arrives and resets the prediction error.

Several SEC techniques have been proposed for block-coded video/images. Many of them are based on some type of interpolation, trying to exploit the correlations between adjacent pixels. In [6], a simple spatial interpolation is used. In [7] a directional extrapolation algorithm was proposed,

J. Koloda, V. Sánchez and A. M. Peinado are with the Department of Signal Theory, Networking and Communications, University of Granada, Granada, Spain (e-mail: janko@ugr.es; victoria@ugr.es; amp@ugr.es).

J. Østergaard and S. H. Jensen are with the Department of Electronic Systems, Aalborg University, Aalborg, Denmark (e-mail: jo@es.aau.dk; shj@es.aau.dk).

This work has been supported by the Spanish MEC/FEDER project TEC 2010-18009.

which exploits the fact that high frequencies, and especially edges, are visually the most relevant features. An algorithm for preservation of edges and borders in the transformed domain based on projections onto convex sets has been also proposed [8]. A technique including edge detectors combined with a Hough transform, a powerful tool for edge description, was utilized in [9]. A more advanced Hough transform based method was proposed in [10]. However, the performance of these methods drops when multiple edges or fine textures are involved. Modelling natural images as Markov random fields for EC was treated in [11]. This scheme produces relatively small squared reconstruction errors at the expense of an over-smoothed (and, therefore, blurred) image. The authors in [12] combined edge recovery and selective directional interpolation in order to achieve a more visually pleasing texture reconstruction. A content adaptive algorithm was introduced in [13]. A simple interpolation is applied if there are only a few edges crossing the missing macroblock and a best-match approach is applied if the macroblock is decided to contain texture. For this algorithm, and in general for all switching SEC techniques, a correct classification is critical since an erroneous decision on the macroblock behaviour could have a very negative effect on the final reconstruction. Inpainting-based methods can also be adopted for SEC purposes [14] [15]. Sequential pixel-wise recovery based on orientation adaptive interpolation is treated in [16]. As we will show later, pixel by pixel recovery usually suffers from smoothing high frequency textures. In [17], Bayesian restoration is combined with DCT pyramid decomposition. Bilateral filtering exploiting a pair of Gaussian kernels is treated in [18]. The algorithm seems quite competitive although some high frequency textures may be found overfiltered. Recently, SEC techniques in transform domains [19] have shown promising results although ringing can be observed in some cases.

TEC techniques take advantage of temporal and/or spatial redundancy as well. A joint video team (JVT) reference software TEC algorithm includes frame copying and motion vector copying [20]. A more advanced recovery of lost motion vectors is based on the boundary matching algorithm (BMA) [21] that minimizes the squared error between the outer boundary of the lost macroblock and the inner boundary of macroblocks found in the reference frame. A slight modification of BMA, overlapping BMA (OBMA), matches the outer boundaries of both the missing macroblock and the reference, leading to more accurate reconstructions [21]. These techniques, however, consider a linear movement and assume that the entire macroblock has been moved the same way. This issue is palliated by a multi-hypothesis approach (based on BMA) [22] which, however, lacks in generality. In [23], MV's are estimated by a Lagrangian interpolation of previously extrapolated MV's. This technique is entirely based on MV's so maintaining spatial continuity may be an issue. An edge-directed hybrid EC algorithm was proposed in [24]. Strong edges are estimated first and regions along these edges are recovered afterwards. Another combined EC technique is presented in [25]. It is a modification of the classic BMA under spatio-temporal constraints with an eventual posterior refinement based on partial differential equations. However, the improvement over the BMA is rather moderate. A MAP estimator, using an adaptive Markov random field process, is used to conceal the lost macroblocks in [26]. A statistically driven technique, based on a Gaussian mixture model is obtained in [27] from spatial and temporal surrounding information. This model, however, requires an extensive offline training. A computationally lighter version is described in [28]. Interesting results are obtained in [29] where a sparse representation based on local dictionaries is used for image reconstruction. This method, however, lacks in flexibility when complex textures are present and the concealment in scanning order may not always be appropriate. Recently, refinement technique [30] based on spatial and temporal AR models has been proposed. However, it is highly dependent on the previous MV estimate (using BMA, for example) and it assumes that (small) groups of macroblocks can be modelled using the same AR process which for low resolution videos or complex scenes may be inaccurate.

In this paper we propose an error concealment technique that automatically adapts itself to SEC [31], TEC or a combined SEC/TEC scheme according to the available information. Our proposal tries to fix or palliate some of the weak points of the previously referenced work such as blurring, blocking or filling order. The lost regions are recovered sequentially using a linear predictor whose coefficients are estimated by an adaptive procedure based on sparsity and a missing data imputation approach. First, we formulate the problem of estimating the predictor coefficients (only for SEC) as a convex optimization problem and then we derive an efficient alternative based on an exponential approximation. Although different exponential estimators have been used in EC algorithms [17] [18], a thorough treatment, combined with a linear prediction model, sparse recovery and sequential filling is proposed in this paper. This leads to a more generic and flexible EC technique. We also show that our EC scheme can be straightforwardly extended to also account for temporal correlations in video sequences (TEC and SEC/TEC). The experimental results show that our proposals provide better performance than other existing state-of-the-art algorithms on a wide selection of images and video sequences. In particular, the exponential approximation provides the best perceptual results.

The paper is organized as follows. In Section II we formulate the problem and introduce the linear prediction image model employed in the optimization process as well as the estimator (linear predictor) used for EC. The convex optimization based error concealment algorithm and its exponential approximation are presented in Section III. The model for video sequences is treated in Section IV. Simulations results and comparisons with other SEC and TEC techniques are presented in Section V. The last section is devoted to conclusions.

II. Linear Prediction Modelling and its Application to Error Concealment

Our aim is to conceal a lost region by optimally exploiting the correlations with the correctly received and decoded pixels in its neighbouring area. These correlations will be modelled

and exploited by means of vector linear prediction as it is described in the next section.

The following subsections describe how this model can be suitably estimated and applied to our concealment task.

A. Vector LP-based spatial modelling

Let us assume that our image can be modelled as a stationary random field. Then, we can expect that every pixel z can be linearly predicted from a small set of surrounding pixels. The corresponding linear prediction (LP) model is defined by,

$$z = \sum_{(k,l)\in\mathcal{R}_z} w(k,l)z(k,l) + \nu \tag{1}$$

where $w(k,l)$ are the LP coefficients, $\mathcal{R}_z$ is the region of surrounding pixels employed for prediction, and ν is the residual error. We will assume integer pixel values belonging to $\Psi = [0, 255]$ for each colour space component.

In our case, we are interested in LP-based reconstruction of groups of lost pixels. Thus, it is convenient to re-formulate the above LP spatial modelling into a vector form by replacing the pixels $z(k,l)$ in (1) by pixel vectors. Let $\boldsymbol{z}$ be an arbitrarily shaped group of pixels that we want to express in terms of our LP model. Writing $\boldsymbol{z}$ as a column vector, we have that $\boldsymbol{z} \in \Psi^n$, where n is the number of pixels contained in $\boldsymbol{z}$. Also, let $\mathcal{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{|\mathcal{Z}|}\}$ be the set of all possible spatially shifted versions of $\boldsymbol{z}$ which are employed to predict it. Then, the whole region employed to predict $\boldsymbol{z}$ is,

$$\mathcal{N}_{\boldsymbol{z}} = \bigcup_{j=1}^{|\mathcal{Z}|} \boldsymbol{z}_j. \tag{2}$$

Again, we can expect that prediction can be carried out with a small number $|\mathcal{Z}|$ of neighbouring vectors. Now, Eq. (1) can be extended to a vector form as follows, [1]

$$\boldsymbol{z} \;=\; \sum_{j=1}^{|\mathcal{Z}|} w_j \boldsymbol{z}_j + \boldsymbol{\nu}, \tag{3}$$

where $\boldsymbol{\nu}$ is the corresponding vector of residuals and $w_j \geq 0$ for all $j = 1, \ldots, |\mathcal{Z}|$.

The previous LP model can be applied to estimate $\boldsymbol{z}$ from the known neighbour vectors in region $\mathcal{N}_{\boldsymbol{z}}$ as,

$$\hat{\boldsymbol{z}} \;=\; \sum_{j=1}^{|\mathcal{Z}|} w_j \boldsymbol{z}_j. \tag{4}$$

In order to obtain optimal LP coefficients, the residual energy

$$\epsilon(\boldsymbol{w}) \triangleq \|\boldsymbol{\nu}\|^2 = \left\| \boldsymbol{z} - \sum_{j=1}^{|\mathcal{Z}|} w_j \boldsymbol{z}_j \right\|_2^2 \tag{5}$$

is usually minimized by solving a system of normal equations.

[1] Note that the intraprediction scheme used the in H.264 codec is a particular case of (3).

B. Application to error concealment: sparse LP

We will denote $\mathcal{S}$ as the set of known pixels and $\mathcal{L}$ will denote the set of lost pixels (see Fig. 1(a)). When applying the above LP estimator of Eq. (4) to compute a lost group of pixels $\boldsymbol{z}$, we are facing two problems:

1) Since $\boldsymbol{z}$ is not known, it is not possible to find the residual energy function $\epsilon(\boldsymbol{w})$ exactly. In order to solve this problem, a solution based on missing-data imputation is proposed later in this section.
2) The region $\mathcal{N}_{\boldsymbol{z}}$ required for prediction is not known either. Instead, we have to employ a support area $\mathcal{S}$ of available (correctly received and decoded) pixels which provides us with a set $\mathcal{Z}'$ containing $M = |\mathcal{Z}'|$ available neighbour vectors $\boldsymbol{z}_j$ ($j = 1, \ldots, M$), that is,

$$\mathcal{S} = \bigcup_{j=1}^{M} \boldsymbol{z}_j. \tag{6}$$

Then, some pixels required for prediction in (4) may be missing. Also, since the image is, in general, non-stationary, the support area $\mathcal{S}$ may include a high number of alien pixels not useful for predicting $\boldsymbol{z}$ ($M \gg |\mathcal{Z}|$, typically). As a result, the usual least-squares solution based on solving a system of normal equations is not suitable in our case. Typically, this solution involves the inversion of a huge $M \times M$ correlation matrix of small rank which would lead us to a poor solution. This small rank indicates that the number of vectors $\boldsymbol{z}_j \in \mathcal{S}$ useful for prediction is quite small. In other words, we can say that the solution $\boldsymbol{w} = (w_1, \ldots, w_M)^t$ we are seeking will be a sparse vector.

In order to overcome this last problem, the classical least-squares estimation of the LP coefficients can be replaced by a joint optimization of the squared error of Eq. (5) and the level of sparsity of the solution (typically represented by the ℓ_0-norm), which leads to a sparse linear prediction (SLP) scheme [32]. This scheme yields an unconstrained minimization problem, that we will represent as the following constrained optimization [33]:

$$\begin{aligned} \text{minimize} \quad & \epsilon(\boldsymbol{w}) = \left\| \boldsymbol{z} - \sum_{j=1}^{M} w_j \boldsymbol{z}_j \right\|_2^2 \\ \text{subject to} \quad & \|\boldsymbol{w}\|_0 \leq \delta_0 \;\text{ and }\; \boldsymbol{w} \succeq 0. \end{aligned} \tag{7}$$

where δ_0 is a parameter that controls the sparsity level and $\boldsymbol{w} \succeq 0$ is imposed to prevent negative pixels from the estimator (9) which is introduced later in this section. Moreover, preliminary experiments have shown that not using this last constraint would yield a worse performance.

This optimization involves two problems that will be addressed in the next section. First, we have that the ℓ_0-norm is non-convex and unfortunately also computationally infeasible for problems of higher dimensions. This problem is usually solved through convex relaxation. Second, we have the problem of selecting a suitable maximum value for sparsity parameter δ_0. We will shortly see that convex relaxation of (7) also provides a natural and smart solution to this issue which is proposed in Section III.

The LP formulation in (7) provides us with an adaptive procedure which dynamically obtains both the LP coefficients and the region of support $\mathcal{N}_{\boldsymbol{z}}$ (defined by those vectors $\boldsymbol{z}_j$ with $\boldsymbol{w}_j \neq 0$) for every image block $\boldsymbol{z}$. We still have the problem of $\boldsymbol{z}$ being unknown. As a consequence, the squared error $\epsilon(\boldsymbol{w})$ cannot be directly computed. In order to solve this, we will adopt a missing data approach where lost pixels can be imputed from known ones [34]. Instead of having a vector $\boldsymbol{z}$ completely unknown, we will consider that it contains both known and unknown pixels. Without loss of generality, let $\boldsymbol{z}$ be a group of pixels as shown in Fig. 1(a). Let the vector $\boldsymbol{z} = \boldsymbol{x} \cup \boldsymbol{y}$ consist of the two subvectors $\boldsymbol{x}$ and $\boldsymbol{y}$, where $\boldsymbol{x}$ denotes the missing pixels and $\boldsymbol{y}$ denotes correctly received and decoded pixels and can be seen as the spatial context of $\boldsymbol{x}$. Every $\boldsymbol{z}_j \in \mathcal{Z}'$ is split in a similar way, as shown in Fig. 1(a). Since $\boldsymbol{z}$ is (locally) stationary and $\boldsymbol{y} \subset \boldsymbol{z}$, then we can approximate the weights obtained from (7) by means of the following procedure:

$$\begin{aligned} \text{minimize} \quad & \epsilon_{\boldsymbol{y}}(\boldsymbol{w}) = \left\| \boldsymbol{y} - \sum_{j=1}^{M} w_j \boldsymbol{y}_j \right\|_2^2 \\ \text{subject to} \quad & \|\boldsymbol{w}\|_0 \leq \delta_0 \ \text{ and } \ \boldsymbol{w} \succeq 0. \end{aligned} \tag{8}$$

Section III will be devoted to the search for solutions to this optimization problem.

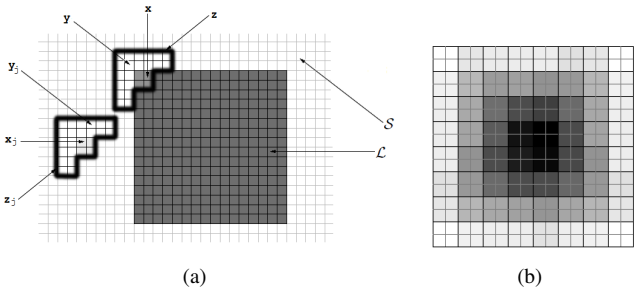
Finally, according to (4) the concealed group of pixels, $\hat{\boldsymbol{x}}$, can be approximated by a linear combination of blocks within its neighbourhood

$$\hat{\boldsymbol{x}} \quad = \quad \sum_{j=1}^{M} w_j^* \boldsymbol{x}_j, \tag{9}$$

where $\boldsymbol{w}^* = (w_1^*, \ldots, w_M^*)^t$ is the vector of optimal weights (LP coefficients) obtained by (8).

C. Application to error concealment: sequential filling

The H.264/AVC encoder packetizes the stream by slices so a loss of one packet implies a loss of, at least, one 16×16 macroblock. Applying (9) to $\boldsymbol{x} \in \Psi^{16\times16}$ would lead to significant imprecisions due to blocking as well as blurring since it is often not possible to find a combination of $\boldsymbol{x}_j$'s suitably matching $\boldsymbol{x}$ due to the high number of dimensions in $\Psi^{16\times16}$. This means that the residual error from (3) may still carry significant energy. This is the reason why the H.264/AVC standard also includes submacroblock prediction [3]. In order to manage with this problem, we introduce sequential recovery. Thus, the macroblock is recovered using a set of square patches $\hat{\boldsymbol{x}} \in \Psi^{p\times p}$ with $1 \leq p \leq 16$. Pixel-wise reconstructions ($p = 1$), as in [16], may introduce considerable blurring when high frequencies are involved (Fig. 11(b)). By using groups of pixels the correlation within a group is better preserved and so is the texture (Fig. 11(c)). Let us consider, without loss of generality, $p = 2$ and let $\boldsymbol{y}$ include all the received and already recovered pixels within the 6×6 block with the lost pixels $\boldsymbol{x}$ placed in its centre, as shown in Fig. 1(a). The macroblock is recovered sequentially by filling it with $\hat{\boldsymbol{x}}$ obtained by applying (8) and (9). The filling order is critical and it should preserve the continuity of image structures [15]. In [15], the filling priorities of every patch are set in order to maintain the continuity of isophotes and according to the amount of information within the patch. Our proposal, due to the shape of the context $\boldsymbol{y}$, can achieve an appropriate filling order in a much simpler way by using contexts reliabilities. We define the reliability ρ of context $\boldsymbol{y}$ as the sum of reliabilities of all its pixels. Initially, the reliability of a pixel is set to 1 if it has been correctly received and decoded. Missing pixels have reliability zero. When a pixel $x \in \boldsymbol{x}$ is concealed, its reliability is set to $\alpha\rho/m$, where $0 < \alpha < 1$ and m is the number of pixels contained in $\boldsymbol{y}$. We use $\alpha = 0.9$ in our simulations. The lost region $\boldsymbol{x}$, whose context $\boldsymbol{y}$ produces the highest reliability, is recovered first. The reliability is non-increasing and the reconstruction evolves from the outer layer towards the centre of the missing macroblock. Figure 1(b) shows the filling order of a 16×16 macroblock using 2×2 patches. Note that the first patches to be concealed are the corners as their contexts are the largest ones, and thereby providing more reliable information (which leads to a more accurate estimate of the LP coefficients).



(a) (b)

Fig. 1. (a) Example of configuration for the vectors $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$. $\mathcal{S}$ denotes the set of known pixels and $\mathcal{L}$ denotes the set of lost pixels. (b) Filling order for sequential reconstruction with 2×2 patches ($p = 2$). The regions illustrated by brighter level are recovered first.

III. LP PARAMETER ESTIMATION

The scheme proposed in the previous section requires the computation of a set of LP coefficients by solving the optimization problem of Eq. (8). In this section, we propose first a solution based on convex relaxation. Then, we derive a computationally less expensive algorithm by applying several approximations.

A. SLP via convex relaxation (SLP-C)

The main problem that arises when solving (8) is that the ℓ_0-norm is non convex, so that this optimization usually requires exhaustive search and is therefore computationally prohibitive. Applying convex relaxation [35], the solution to the optimization defined by (8) can be modified in terms of the ℓ_1-norm as follows:

$$\begin{aligned} \text{minimize} \quad & \epsilon_{\boldsymbol{y}}(\boldsymbol{w}) = \left\| \boldsymbol{y} - \sum_{j=1}^{M} w_j \boldsymbol{y}_j \right\|_2^2 \\ \text{subject to} \quad & \|\boldsymbol{w}\|_1 \leq \delta_1 \ \text{ and } \ \boldsymbol{w} \succeq 0. \end{aligned} \tag{10}$$

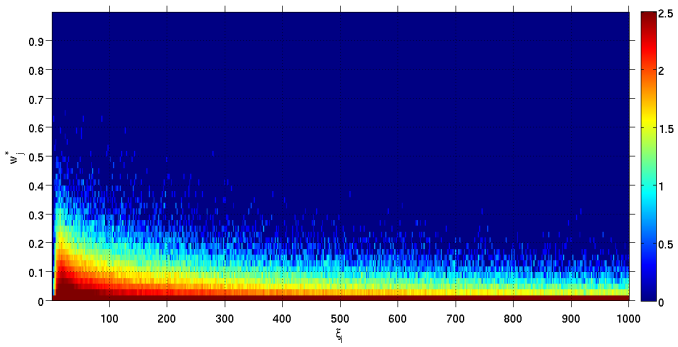


Fig. 2. Histogram of pairs squared-error/weight (ξ_j, w_j^*) for *Lena*. Logarithmic scale is employed for more clarity. For reconstruction purposes 2×2 patches are used and loss pattern from Fig. 6(b) is applied.

In our simulations, this optimization is solved by the primal-dual interior point (IP) method [36].

The remaining problem is the selection of a suitable sparsity level δ_1 (redefined under the ℓ_1-norm). In order to do this, we will assume smoothness in the visual features of an image. This implies that the reconstructed block should not contain any singular features. In the particular case of luma, it means that a reconstructed pixel could not be brighter (darker) than the brightest (darkest) pixel in $\mathcal{S}$. This requires that (9) must be a convex combination and it implies that $\delta_1 = 1$. The resulting technique will be referred to as SLP-C in the following.

B. SLP with exponentially distributed weights (SLP-E)

Although there are efficient algorithms for solving convex optimization problems, such as the IP method employed above, the processing time still remains very high and far from real-time. In this section we develop a fast approximation for solving the minimization problem in (10). Specifically, we show that the optimal weights $\boldsymbol{w}^*$ obtained from (10) can be well modelled by an exponential function.

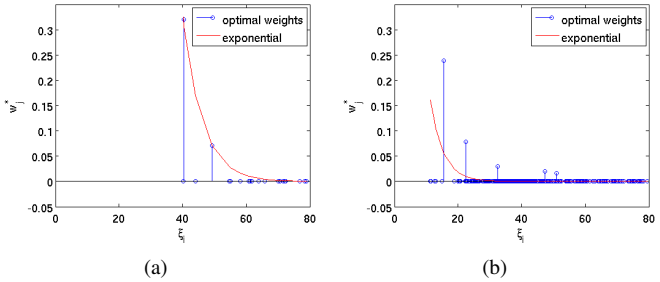
According to (10), every context $\boldsymbol{y}_j$ has a weight w_j^* associated. Due to the high spatial correlation of an image, it is likely that contexts that produce smaller squared error, ξ_j, would generate larger weights, where we define the squared error ξ_j associated to a context $\boldsymbol{y}_j$ as,

$$\xi_j = \frac{\|\boldsymbol{y} - \boldsymbol{y}_j\|_2^2}{m}. \tag{11}$$

Figure 2 represents the joint 2D histogram of pairs (ξ_j, w_j^*) for the image of *Lena*. The loss pattern applied is the one shown in Fig. 6(b). The histogram suggests that there is an exponential relationship between the squared errors ξ_j and the weights w_j^*. With this in mind, we propose the following approximation for the LP weights:

$$\hat{w}_j = C\exp\left(-\frac{1}{2}\frac{\xi_j}{\sigma^2}\right), \tag{12}$$

where σ^2 is a decay factor that controls the slope of the exponential and C is a normalization factor that ensures the



(a) (b)

Fig. 3. Example of the exponential estimated by means of the optimal weights $\boldsymbol{w}^*$ for two different patches.

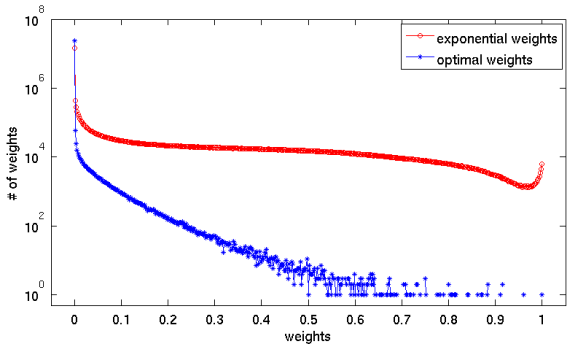


Fig. 4. Comparison of the weights histograms obtained by SLP-E (red) and SLP-C (blue) for the image of *Lena*. The vertical axis uses a logarithmic scale for a clearer visualization and σ^2 has been fixed to 10 for the whole image.

sparsity constraint $\|\boldsymbol{w}\|_1 = 1$, that is,

$$C = \frac{1}{\sum_{i=1}^{M} \exp\left(-\frac{1}{2}\frac{\xi_i}{\sigma^2}\right)}. \tag{13}$$

Note that this normalization always forces the solution $\hat{\boldsymbol{w}}$ to have the maximum value of sparsity considered in (10), i.e. $\delta_1 = 1$. The corresponding LP estimator is obtained by replacing the optimal weights w_j^* by their exponential approximation $\hat{w}_j$ in Eq. (9). The resulting EC technique will be referred to as SLP-E in the following.

Let us analyze the approximation proposed in equations (12) and (13). We can see that the exponential trend observed in Fig. 2 cannot be written down as a single exponential function for the whole image. In fact, the figure shows lots of exponential contours. There are two reasons for this:

1) We must take into account the effect of the mild sparsity constraint applied in (10). Thus, given several similar contexts $\boldsymbol{y}_j$ (representing a certain context type) with small quadratic errors ξ_j (that is, relevant for reconstruction), the optimization algorithm picks one context and suppresses the others instead of using all of them. On the contrary, the exponential approximation relaxes the sparsity constraint and keeps all the relevant contexts.
2) We must also consider that Fig. 2 shows all the pairs (ξ_j, w_j^*) for all the patch linear predictors in the image. However, clearly all these linear predictors are different and must have a different factor σ^2, since this is the only free parameter in Eq. (12).

(a) (b) (c)

Fig. 7. EC with SLP-E for different values of σ^2: (a) σ^2 = 0.5, numerically unstable reconstructions are represented with white level, (b) σ^2 = 10, (c) σ^2 = 50.

Let us consider first the issue of obtaining a suitable value of σ^2 for each patch predictor. This factor is related to the squared error ϵ_y and, therefore, to the local predictability of the image signal. In order to estimate a suitable value of σ^2 for every predictor, a logical solution is that of minimizing the prediction error $\epsilon_{\boldsymbol{y}} = \epsilon_{\boldsymbol{y}}(\sigma^2)$ defined in (10) but constrained to the LP weights defined by Eqs. (12) and (13). Figure 3 illustrates two examples of the optimal weights and their corresponding exponential approximations with factors σ^2 estimated as described above. In the first example, the exponential function mainly follows the most relevant optimal weights. However, in the second one, the exponential approximation leads to weights which are smaller than the optimal ones. In order to understand this, we must take into account that there is a considerable number of zero-valued optimal weights in the small squared error area, which is due, as previously explained, to the mild sparsity constraint. On the contrary, the exponential approximation introduces a sparsity relaxation and the weight assigned to a certain type of context is distributed among the contexts of that type through the normalization in (13) and the selection of a suitable σ^2. We must point out that the sparsity relaxation just described is quite limited. In order to see this, the histograms for both optimal and exponential weights are depicted in Fig. 4. We can see that although the exponential approximation reduces sparsity, most of the weights are still close to zero.

Table I shows the mean value and the standard deviation of σ^2 for several tested images. $\epsilon_{\boldsymbol{y}}(\sigma^2)$ minima have been obtained by exhaustive search. In the following and for the sake of computational simplicity, a fixed value of σ^2 will be used. Simulations reveal that this simplification, along with the exponential approximation, leads to a factor of 100 of computational saving with respect to SLP-C. For natural images, σ^2 values around 10 lead to visually good results (Fig. 7(b)). Larger values of σ^2 may lead to oversmoothing (Fig. 7(c)) while smaller values may lead to numerical instability and should be avoided (Fig. 7(a)) (unless the image is extremely stationary).

Finally, we must also point out that the approach developed here can be alternatively interpreted as a non-parametric kernel-based regression, in particular, as a Nadaraya-Watson estimator.

σ^2	*Lena*	*Clown*	*Office*	*Barbara*	Average
mean	6.70	12.01	7.35	12.99	9.76
std	14.69	35.53	15.50	20.09	21.45

TABLE I
ESTIMATED VARIANCE (MEAN VALUE AND STANDARD DEVIATION) FOR TESTED IMAGES.

IV. TEMPORAL MODEL OF A VIDEO SEQUENCE

The importance of temporal correlations is reflected by the fact that they are a crucial issue in video coding. However, in the temporal domain, video signals tend to be non-stationary due to motion. That is, the pixel $z(i,j)$ in the current frame usually cannot be predicted using the pixels with the same location in previous frames [37]. This can be palliated by applying motion compensation. In fact, the H.264/AVC standard encodes the submacroblock $sMB^{(n)}_{(i,j)}$ belonging to the current P-frame n as

$$sMB^{(n)}_{(i,j)} \quad = \quad sMB^{(n-\tau)}_{(i+MV(i),j+MV(j))} + \boldsymbol{r} \qquad (14)$$

where $MV(i,j)$ is the motion vector, $\boldsymbol{r}$ is the residual error and τ is the temporal lag to the reference frame n. Note that τ depends on visual properties of the video as well as the dimension of the prediction buffer. Moreover, regardless of the buffer size, the encoder selects the sparsest set of weights since only one reference submacroblock is taken into account. For B-frames and P-frames where weighted prediction is applied, two reference submacroblocks are utilized.

The estimation scheme of Section II can be straightforwardly extended in order to account for both temporal and spatial correlations. In this case, Eq. (3) could be seen as a generalization of (14). The stationary region $\mathcal{N}_{\boldsymbol{z}}$ will now not only comprise pixels from the current frame but also pixels from the previous frames. As in the case of SEC, the stationary 3D region is unknown and the whole support area $\mathcal{S}$ needs to be searched. We will set the support area to include all the available neighbouring macroblocks from the current frame (as in the previous section) and all the corresponding macroblocks from the previous frame. For the sequences of *Foreman* and *Stefan*, more than 99% of MV have $\tau = 1$, so considering only the previous frame is a reasonable simplification. Figure 8 illustrates an example where the corrupted frame utilizes dispersed slicing and the previous frame is received without errors.

In practice, the loss of a NALU implies that residual errors as well as motion vectors are lost (unless data partitioning is applied at the encoder side at the expense of a higher bit-rate) [3]. In order to obtain high quality predictions, the support area $\mathcal{S}$ should include all the motion compensated pixels located within the corrupt macroblock. For a standard frame rate of 30 fps, the motion vectors between two consecutive frames are likely to be moderate. In fact, Fig. 9 shows the histograms of motion vectors norm for four different 30-frame video sequences. It follows from the histograms that the support area composed as described above covers more than 95% of motion vectors. In other words, in less than 5% of cases the

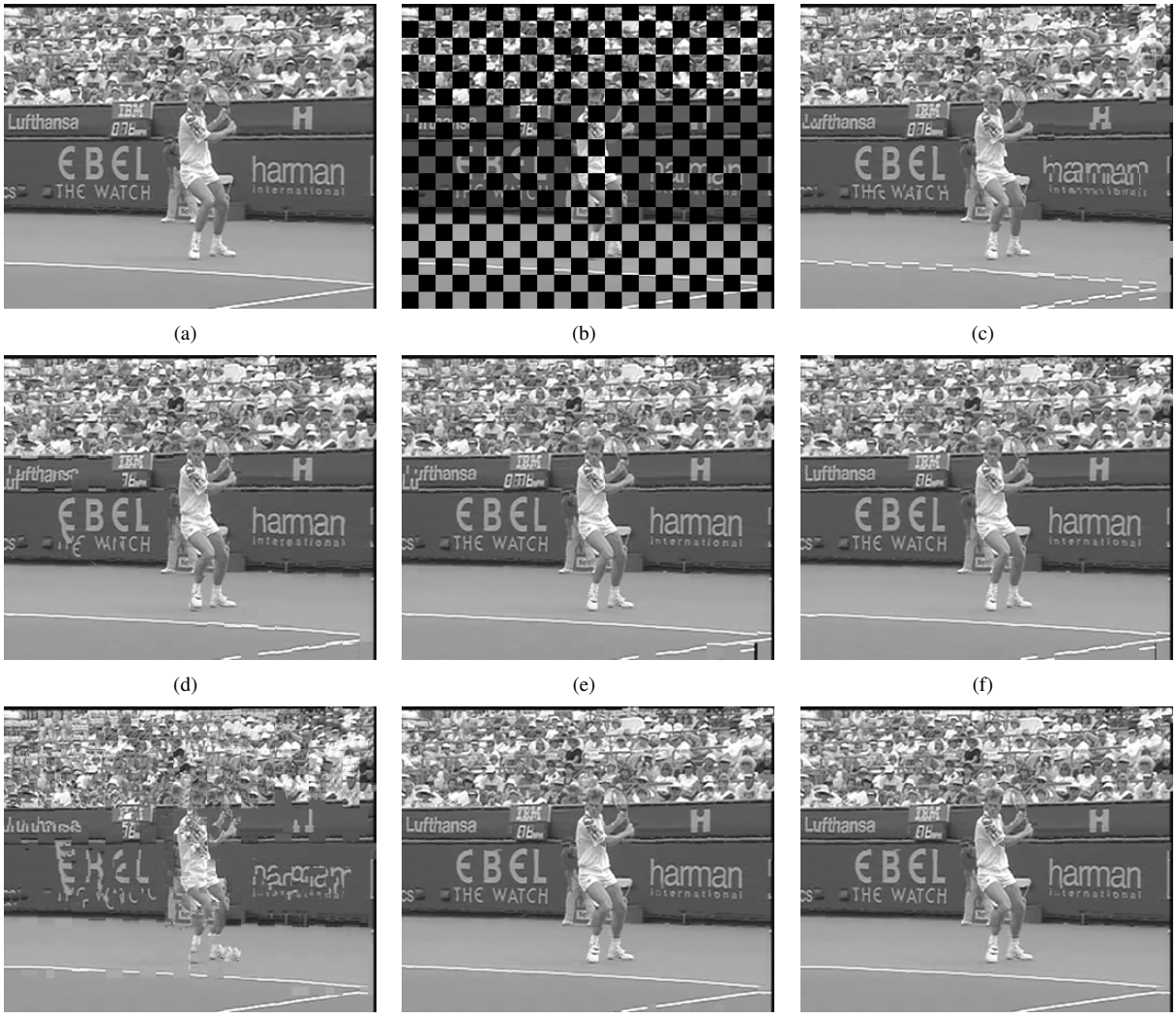


Fig. 5. Comparison of reconstructions obtained by different procedures. (a) Original frame. (b) Received frame. (c) Reconstruction by average MV replacement (PSNR = 21.36, MS-SSIM = 86.52). (d) Reconstruction by BMA (PSNR = 22.44, MS-SSIM = 90.24). (e) Reconstruction by OBMA (PSNR = 24.14, MS-SSIM = 96.86). (f) Reconstruction by MHEC (PSNR = 24.37, MS-SSIM = 96.93). (g) Reconstruction by SLP-E using spatial information only (PSNR = 18.56, MS-SSIM = 77.80). (h) Reconstruction by SLP-E using temporal information only (PSNR = 25.79, MS-SSIM = 97.73). (i) Reconstruction by SLP-E using both spatial and temporal information (PSNR = 25.93, MS-SSIM = 97.74).

motion compensated macroblock lies (completely or partially) outside the support area (MV amplitude greater than 16). For the sake of computational simplicity, we assumed that the motion vectors were calculated using only the previous frame. The more motion vectors that are covered, the better reconstructions would be obtained as a more complete set of motion compensated pixels (useful for prediction) is used. However, the processing time increases with $|\mathcal{S}|$ so applying the proposed support area is a reasonable trade-off. Using this support area, the weights will be computed in the same way as in (12).

Note that pixels from the surroundings (within the current frame) of the missing macroblock are also included. Thus, the algorithm automatically decides whether to use SEC, TEC or combined concealment. This is the consequence of dynamically obtaining the LP coefficients and estimates the

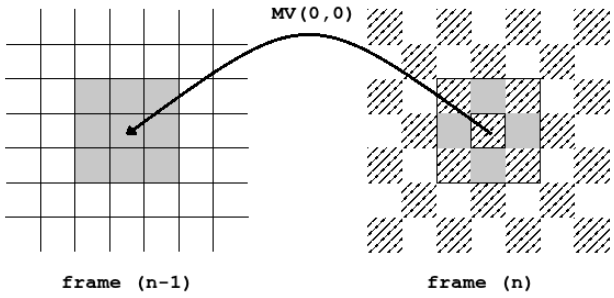


Fig. 8. Support area $\mathcal{S}$ (grey 16×16 macroblocks) for combined TEC/SEC. The striped macroblocks are lost.

stationary area $\mathcal{N}_z$, as discussed in Section II-B. For example, if the previous frame belongs to a different scene, all relevant

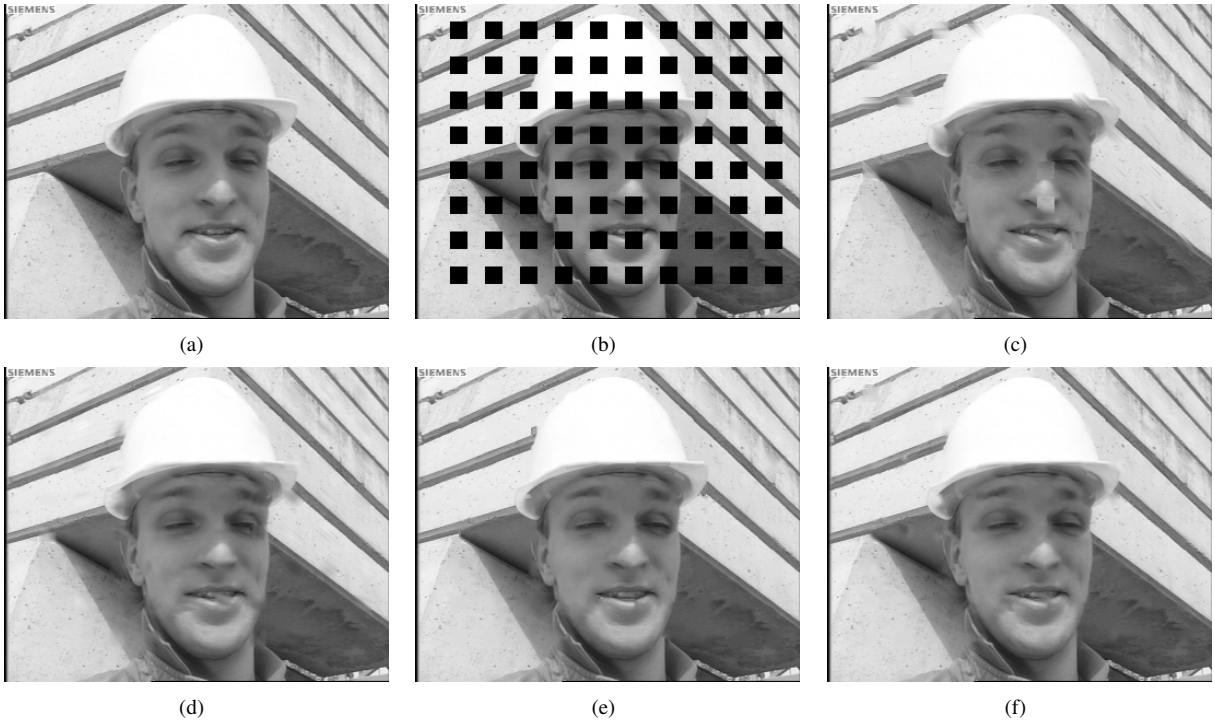


Fig. 6. SEC for the image of *Foreman* (a) Original image, (b) Received data, (c) Reconstruction using CAD (PSNR = 31.46dB, MS-SSIM = 97.54), (d) FSE (PSNR = 34.17dB, MS-SSIM = 98.03), (e) SLP-E (PSNR = 35.46dB, MS-SSIM = 98.73), (f) SLP-C (PSNR = 35.48dB, MS-SSIM = 98.68).

PSNR	BIL	POC	EXT	SHT	CAD	AVC	MRF	INP	BLF	OAI	FSE	SLP-E	SLP-C	ORA
Average	26.92	25.94	26.76	27.19	29.12	28.08	29.12	28.78	29.71	30.15	30.20	30.32	**30.90**	31.36
Lena	30.00	28.04	29.39	30.47	30.44	30.42	32.17	30.88	32.15	32.82	32.72	32.55	**32.85**	33.41
Goldhill	30.00	28.50	29.57	29.97	30.24	31.27	31.12	30.40	30.91	31.54	31.78	31.54	**32.07**	32.97
Foreman	27.12	28.49	29.26	28.34	31.46	29.11	32.98	33.87	34.75	35.03	34.18	35.46	**35.48**	37.38
Barbara	26.19	24.30	25.85	26.40	26.78	26.85	27.99	28.04	29.91	29.66	30.84	30.79	**31.91**	32.15
Office	27.54	27.56	27.32	27.54	29.43	29.99	29.77	29.64	30.06	31.77	31.33	31.30	**32.06**	32.68
Cameraman	25.96	23.66	24.82	26.16	26.51	26.14	26.67	25.45	26.03	27.27	**27.44**	27.24	27.27	27.28
Baboon	24.15	24.63	24.72	24.14	24.92	25.42	26.14	25.06	26.05	26.06	26.02	25.70	**26.21**	25.93
Clown	27.76	24.36	26.30	27.62	29.12	28.55	28.23	27.89	28.73	29.75	29.19	27.39	**30.79**	31.00
Tire	23.59	23.92	23.82	24.10	24.47	25.43	27.00	26.37	28.76	27.42	28.31	28.77	**29.32**	29.43
MS-SSIM														
Average	93.83	91.23	93.82	94.12	95.69	94.74	95.87	95.53	96.35	95.81	96.36	**97.04**	96.58	97.80
Lena	96.72	92.85	96.52	97.03	96.59	96.56	97.64	96.65	97.44	97.65	97.80	**97.97**	97.75	98.48
Goldhill	93.83	92.50	94.81	93.86	94.53	95.65	95.71	95.21	95.52	95.62	96.14	**96.43**	96.34	97.50
Foreman	95.16	93.09	97.16	95.65	97.58	96.87	98.10	98.22	98.15	98.68	97.92	**98.70**	98.65	99.19
Barbara	95.24	89.42	94.70	95.57	95.73	94.87	96.00	95.72	97.04	97.07	97.64	97.92	**98.12**	98.67
Office	93.93	93.24	94.84	93.92	95.66	95.77	96.21	96.35	96.12	97.27	96.90	**97.45**	97.39	98.32
Cameraman	93.38	87.22	93.14	93.47	94.87	93.72	94.95	93.96	95.97	93.87	94.31	**96.55**	92.93	96.83
Baboon	88.96	90.16	91.38	88.81	91.89	91.91	93.09	90.95	93.33	92.61	93.32	**93.42**	93.38	94.83
Clown	95.61	91.17	94.09	95.53	96.00	95.55	95.55	95.74	96.22	95.28	96.40	**97.19**	97.13	98.23
Tire	91.65	91.40	93.17	93.12	88.56	92.07	95.59	94.19	97.35	94.26	96.81	**97.70**	97.30	98.16

TABLE II
PSNR VALUES (IN DB) AND MS-SSIM INDICES (SCALED BY 100) FOR TEST IMAGES RECONSTRUCTED BY SEVERAL ALGORITHMS FOR BLOCK DIMENSIONS 16×16. THE BEST PERFORMANCES FOR EACH IMAGE ARE IN BOLD FACE.

weights calculated by (12) will most likely come from the current frame and the contribution of pixels from the previous (uncorrelated) frame will be negligible. Nevertheless, temporal correlation is usually higher than the spatial one and this phenomenon is observed in the reconstruction process. Figure 10 shows the average weight associated with each pixel within the support area for two different video sequences. We see that the contribution of pixels belonging to the previous frame is considerably higher than the contribution of those within the current frame. Simulations show that for standard video test

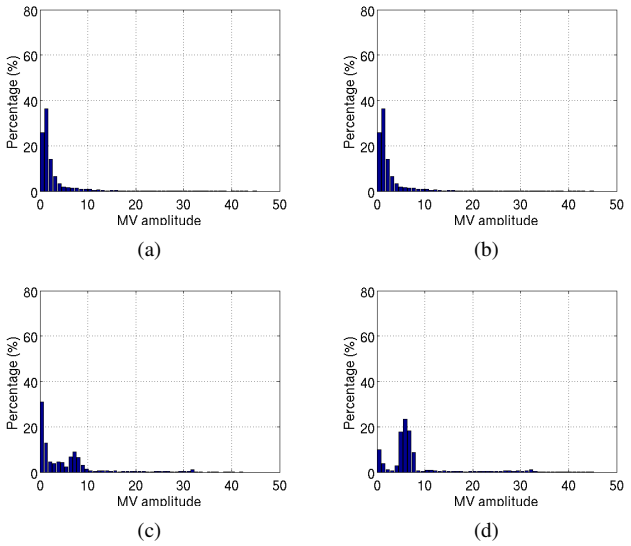


Fig. 9. Histogram of MV amplitudes for the video sequences of (a) *Foreman*, (b) *News*, (c) *Stefan* and (d) *Bus*. Motion vectors were obtained by minimizing the residual error and applying full range search.

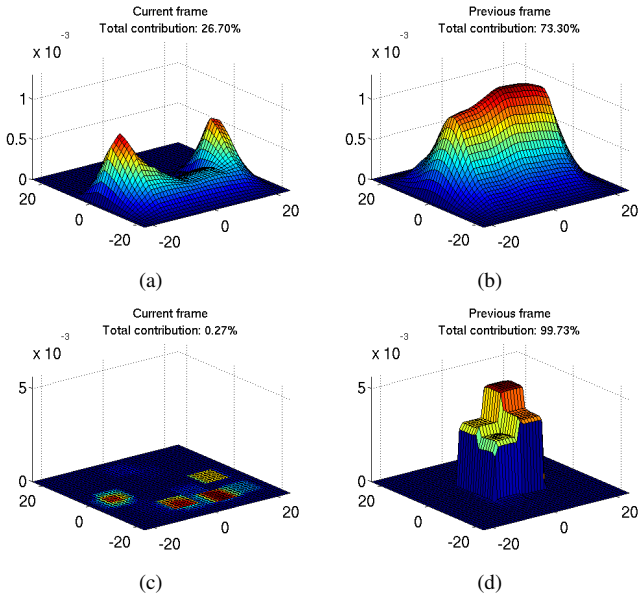


Fig. 10. Average weight per pixel from $\mathcal{S}$ for the sequence of *Stefan* (a)-(b) and *Waterfall* (c)-(d). The percentage indicates the total contribution from pixels from the current frame ((a) and (c)) and from the previous frame ((b) and (d)) to the final reconstruction.

samples, composed by a single shot, the amount of information (pixels) gathered from the previous frame is higher than 70%. Moreover, in some particular cases there will be almost no good template matches within the current frame, as shown in Fig. 10(c) and 10(d). Unlike the pure spatio-temporal hybrid algorithms, our proposal is applicable both for still images (or I-frames) and video. Since temporal correlations tend to be higher than spatial correlations, then smaller values of σ^2 are preferred. Moreover, due to the same reason, larger patches may be utilized to speed up the algorithm and obtain higher quality reconstructions. Here, σ^2 is set to 5 and 8×8 patches are employed.

Figure 5 shows a comparison of our proposal using only spatial information, temporal information and a combination of both with other techniques. In fact, it is observed, at both objective and subjective levels, that using only spatial information achieves poorer quality. The improvement of the combined method over the pure TEC is small, as can be also deduced from Fig. 10. Nevertheless, including spatial information may provide a noticeable visual improvement as can be observed comparing Fig. 5(h) and 5(i).

V. Simulation Results

In order to better take into account the perceptual quality, the multi scale structural similarity (MS-SSIM) index [38] is used for comparison along with the PSNR measure. In the former case, the image is sequentially low-pass filtered and subsampled, so a set of images is obtained, including the original resolution. Then, the SSIM index is applied for every subimage within the set. The SSIM index aims at approximating the human visual system (HVS) response looking for similarities in luminance, contrast, and structure [39]. This index can be seen as a convolution of a fixed-sized mask with the residual error between the reference image and the concealed image [40]. A unique mask size is used for each

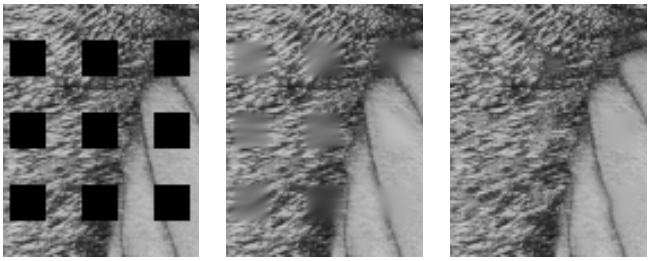


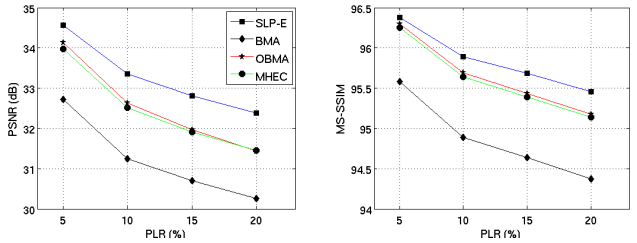
Fig. 11. Example of PSNR and MS-SSIM response to different image reconstructions. (a) Received image, (b) reconstructed by orientation adaptive interpolation (OAI) [16] (PSNR = 27,22, MS-SSIM = 92,47) (c) reconstructed using SLP-E (12) with $p = 2$ (PSNR = 25,56, MS-SSIM = 94,76).

of the images within the set. Therefore fine as well as coarse textures and objects are taken into account.

As shown in Fig. 11, the PSNR does not respond to perceptual visual quality as well as the MS-SSIM index does, since PSNR is a quality criterion merely based on the mean squared error. In spite of that, the weights $\boldsymbol{w}^*$ are obtained according to the squared error (12) since the SSIM index tends to marginalize the influence of changes in intensity [41]. This is a desirable behaviour when measuring the overall perceptual image quality but not when computing predictor coefficients. Thus, the squared error is used when computing the weights while the MS-SSIM index is preferred for an overall quality measure. [2]

The performance of our proposals in SEC mode is tested on the images of *Lena* (512 × 512), *Barbara* (512 × 512), *Baboon* (512 × 512), *Goldhill* (576 × 720), *Clown* (512 × 512), Matlab built-in images *Cameraman* (256 × 256), *Office*

[2]Note that the MS-SSIM index lies in [-1; 1]. In this section, we have scaled the index by 100 in order to better illustrate the differences.



(a) (b)

Fig. 12. Average PSNR (a) and MS-SSIM (b) values vs. packet loss-rates averaged for all the tested video sequences. Tested procedures: SLP-E in the combined SEC/TEC mode, BMA, OBMA and MHEC.

(592 × 896), *Tire* (192 × 224) and the first frame of *Foreman* (288 × 352) sequence. The test is carried out for 16×16 macroblocks and the rate of block loss is approximately 25%, corresponding to a single packet loss of a frame with dispersed slicing structure. We compare the performance with other SEC methods such as bilinear interpolation (BIL) [6], projections onto convex sets (POC) [8], directional extrapolation (EXT) [7], a Hough transform based SEC (SHT) [10], content adaptive technique (CAD) [13], non-normative SEC for H.264 (AVC) [42], Markov random fields approach (MRF) [11], inpainting (INP) [15], bilateral filtering (BLF) [18], frequency selective extrapolation (FSE) [19] and orientation adaptive interpolation (OAI) [16]. [3] Both SLP via convex relaxation (SLP-C) and SLP with exponentially distributed weights (SLP-E) are tested. In the simulations, σ^2 is set to 10 and grey level images are used. Note that a pixel reconstructed by any of the aforementioned algorithms is usually real-valued and does not necessarily belong to Ψ. Thus, reconstructed pixels are rounded to the closest member of Ψ. A subjective comparison of the different algorithms is shown in Fig. 6. As can be seen in Table II, SLP-C provides the best PSNR results as expected, but SLP-E outperforms all the other technique for all the tested images in terms of MS-SSIM, leading so to higher perceptual quality reconstructions. Moreover, the average MS-SSIM and PSNR are superior to those of state-of-the-art algorithms. In addition, an oracle SPL-E (ORA) is included, where the best σ^2 (the value which provides the best reconstruction) is applied for every patch, and it represents the superior limit of the SPL-E performance.

The proposed SLP-E technique in the combined SEC/TEC mode is tested for H.264 coded video sequences of *Foreman*, *Stefan*, *Ice*, *Football*, *Bus*, *Irene*, *Flower* and *Highway*. All sequences employ the common intermediate format (CIF, 352 × 288) and they comprise 30 frames, where only the first frame is intracoded and the remaining frames are predictive coded. An aggressive block loss-rate is applied by utilizing a dispersed slicing structure with two slices per frame (the so-called chessboard structure, see Fig. 5(b)). In this scenario, a loss of one packet implies a loss of 50% of the macroblocks within a frame. However, note that our proposal can be easily extended to other slicing modes. The quantization parameter is set to 25 and the prediction buffer is one frame deep.

[3]Implementations of most of these techniques, as well as the implementation of our algorithm, is available online at [43].

Method	PSNR				MS-SSIM			
	PLR							
	5%	10%	15%	20%	5%	10%	15%	20%
	Foreman							
BMA	35.56	34.03	33.60	32.82	97.35	96.88	96.73	96.48
OBMA	37.69	36.33	35.61	34.90	97.81	97.49	97.30	97.13
MHEC	37.51	36.25	35.70	35.03	97.79	97.51	97.36	97.20
SLP-E	**38.12**	**37.33**	**37.05**	**36.53**	**97.82**	**97.67**	**97.63**	**97.52**
	Stefan							
BMA	29.59	28.18	28.01	27.37	93.84	92.89	92.80	92.37
OBMA	30.46	29.15	28.88	28.31	94.79	94.08	93.96	93.62
MHEC	30.61	29.25	28.98	28.42	94.87	94.14	94.05	93.65
SLP-E	**31.27**	**30.31**	**29.96**	**29.36**	**95.10**	**94.56**	**94.40**	**94.03**
	Football							
BMA	30.37	28.84	28.64	27.95	93.31	92.03	91.75	91.30
OBMA	31.07	29.34	28.93	28.16	93.86	92.44	91.91	91.29
MHEC	30.76	29.15	28.70	28.03	93.60	92.18	91.59	91.14
SLP-E	**31.53**	**30.06**	**29.79**	**29.08**	**94.11**	**93.04**	**92.71**	**92.27**
	Ice							
BMA	34.54	32.39	31.19	31.10	97.76	97.31	97.01	96.92
OBMA	35.25	32.85	31.55	31.16	98.09	97.57	97.23	97.04
MHEC	34.74	32.61	31.24	31.00	97.97	97.45	97.07	96.88
SLP-E	**35.48**	**33.57**	**32.53**	**32.40**	**98.11**	**97.76**	**97.52**	**97.44**

TABLE III
AVERAGE PSNR AND MS-SSIM VALUES FOR DIFFERENT PLR FOR VIDEO SEQUENCES OF *Foreman*, *Stefan*, *Football* AND *Ice*. TESTED PROCEDURES: BMA, OBMA, MHEC AND SLP-E. THE BEST PERFORMANCES FOR EACH SEQUENCE ARE IN BOLD FACE.

Sequence	[24]	[15]	SLP-E	BMA
Foreman	13.28	13.65	9.24	1.00
Irene	9.71	13.13	9.27	1.00

TABLE IV
AVERAGE ERROR CONCEALMENT TIME FOR A CORRUPTED FRAME COMPARED TO BMA.

Packet losses are randomly generated at rates of 5%, 10%, 15% and 20%. For each packet loss rate (PLR), the sequence is transmitted 20 times and the average PSNR and MS-SSIM values are calculated. The proposed technique is compared with other TEC algorithms, namely BMA [21], OBMA [21] and multi-hypothesis EC (MHEC) [22]. The search range for BMA, OBMA and MHEC is [-16, 16] using the zero MV as the starting point, i.e. BMA, OBMA, MHEC and our proposal all work with the same information gathered from the previous frame. The proposed SLP-E outperforms the other techniques for all the tested sequences both in terms of PSNR and MS-SSIM. The results for half of the eight sequences are shown in Table III. PSNR and MS-SSIM values, averaged over all the tested sequences, are shown in Fig. 12. Finally, a subjective comparison is shown in Fig. 5.

Regarding the computational complexity, Table IV shows the processing time ratios of [24], [15] and SLP-E to BMA. We can observe that our proposal requires less processing time than some of the state-of-the-art techniques. Moreover, the average gains of [24] over BMA are approximately 2dB for *Foreman* and 1dB for *Irene* and it outperforms [15] for both cases. Utilizing the same simulation setup as in [24] (dispersed slicing, quantization parameter set to 25 and PLR of 3%, 5%, 10% and 20%), SLP-E achieves average gains over BMA of 2.55dB and 1.20dB, respectively. Thus, SLP-E outperforms both [24] and [15] with less computational burden. Due to the nature of our algorithm, the processing time per MB is approximately constant regardless of the sequence and its resolution, as has been confirmed by the simulations.

Finally, given a multi-scene sequence, the error may occur in the border frame (usually intracoded). In such a case, MV

based techniques fail since they try to extract the concealment information from the previous, and therefore uncorrelated, frame. Modified BMA and OBMA are able to gather the information from the current frame although the reconstructions tend to be of poor quality since both algorithms seek the best match for the entire missing macroblock and this approach usually does not lead to the lower residual energy. Note that the H.264/AVC codec overcome this problem by allowing submacroblock prediction. Moreover, OBMA cannot be applied for all the slicing modes, e.g. this method is unable to conceal the chessboard loss pattern utilizing only the spatial information. On the contrary, due to the sequential filling and the dynamic adaptation to the available information, none of the aforementioned scenarios is an issue for our proposal in the combined SEC/TEC mode.

VI. Conclusions

We have developed a sparse linear prediction estimator, which recovers lost regions in images by filling them sequentially with a weighted combination of patches that are extracted from the available neighbourhood. The weights are obtained by solving a convex optimization problem that arises from a spatial image model. Moreover, we show that the weights can be approximated by an exponential function, so that the resulting method can be alternatively interpreted as a kernel-based Nadaraya-Watson regression. The proposed techniques automatically adapt themselves to SEC, TEC or a combined scenario and can be thus successfully applied to both still images and video sequences.

Our proposals achieve better PSNR and perceptual reconstruction quality than other state-of-the-art techniques. SLP-C is optimized for squared error so it achieves better PSNR than the approximated method. Simulations reveal, however, that SLP-E provides better MS-SSIM. Finally, by applying the approximated algorithm SLP-E the processing time is reduced in a factor of 100.

References

[1] ITU-T, *ITU-T Recommendation H.264*, International Telecommunication Union, 2005.

[2] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, pp. 560–576, July 2003.

[3] I. Richardson, *The H.264 advanced video compression standard*. Wiley, 2010.

[4] A. M. Gomez, A. Peinado, V. Sanchez, and A. Rubio, "Combining media specific FEC and error concealment for robust distributed speech recognition over loss-prone packet channels," *IEEE Transactions on Multimedia*, vol. 8, pp. 1228–1238, November 2006.

[5] A. Peinado, V. Sánchez, and A. Gómez, "Error concealment based on MMSE estimation for multimedia wireless and IP applications," in *Proceedings of PIMRC*, September 2008.

[6] P. Salama, N. Shroff, E. Coyle, and E. Delp, "Error concealment techniques for encoded video streams," in *Proceedings of ICIP*, 1995, pp. 9–12.

[7] Y. Zhao, H. Chen, X. Chi, and J. Jin, "Spatial error concealment using directional extrapolation," in *Proceedings of DICTA*, 2005, pp. 278–283.

[8] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Transactions on Image Processing*, vol. 4, no. 4, April 1995.

[9] D. Robie and R. Mersereau, "The use of Hough transforms in spatial error concealment," in *Proceedings of ICASSP*, vol. 4, 2000, pp. 2131–2134.

[10] H. Gharavi and S. Gao, "Spatial interpolation algorithm for error concealment," in *Proceedings of ICASSP*, April 2008, pp. 1153–1156.

[11] S. Shirani, F. Kossentini, and R. Ward, "An adaptive Markov random field based error concealment method for video communication in error prone environment," in *Proceedings of ICIP*, vol. 6, 1999, pp. 3117–3120.

[12] W. Kung, C. Kim, and C. Kuo, "Spatial and temporal error concealment techniques for video transmission over noisy channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, pp. 789–802, July 2006.

[13] Z. Rongfu, Z. Yuanhua, and H. Xiaodong, "Content-adaptive spatial error concealment for video communication," *IEEE Transactions on Consumer Electronics*, vol. 50, pp. 335–341, February 2004.

[14] P. Harrison, "Texture synthesis, texture transfer and plausible restoration," *PhD. Thesis, Monash University*, 2005.

[15] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on Image Processing*, vol. 13, pp. 1200–1212, September 2004.

[16] X. Li and M. Orchard, "Novel sequential error-concealment techniques using orientation adaptive interpolation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 857–864, October 2002.

[17] G. Zhai, X. Yang, W. Lin, and W. Zhang, "Bayesian error concealment with DCT pyramid for images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 1224–1232, September 2010.

[18] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Image error-concealment via block-based bilateral filtering," in *Proceedings of ICME*, June 2008, pp. 621–624.

[19] J. Seiler and A. Kaup, "Fast orthogonality deficiency compensation for improved frequency selective image extrapolation," in *Proceedings of ICASSP*, March 2008, pp. 781–784.

[20] J. Model, "The JVT reference software for H.264/AVC [online]," *Available: http://iphome.hhi.de/suehring/tml/download*.

[21] T. Thaipanich, P. Wu, and C. Jay Kuo, "Video error concealment with outer and inner boundary matching algorithms," in *Proceedings of SPIE*, 2007.

[22] K. Song, T. Chung, C.-S. Kim, Y.-O. Park, Y. Kim, Y. Joo, and Y. Oh, "Efficient multi-hypothesis error concealment technique for H.264," in *Proceedings of ISCAS*, May 2007, pp. 973–976.

[23] J. Zhou, B. Yan, and H. Gharavi, "Efficient motion vector interpolation for error concealment of H.264/AVC," *IEEE Transactions on Broadcasting*, vol. 57, pp. 75–80, March 2011.

[24] M. Ma, O. Au, S. G. Chan, and M. Sun, "Edge-directed error concealment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, pp. 382–394, March 2010.

[25] Y. Chen, Y. hu, O. Au, H. Li, and C. Chen, "Video error concealment using spatio-temporal boundary matching and partial differential equation," *IEEE Transactions on Multimedia*, vol. 10, pp. 2–11, January 2008.

[26] S. Shirani, F. Kossentini, and R. Ward, "A concealment method for video communications in an error-prone environment," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, June 2000.

[27] D. Persson, T. Eriksson, and P. Hedelin, "Packet video error cocnealment with gaussian mixture models," *IEEE Transactions on Image Processing*, vol. 17, pp. 145–154, 2008.

[28] D. Persson and T. Eriksson, "Mixture model- and least squares-based packet video error concealment," *IEEE Transactions on Image Processing*, vol. 18, pp. 1048–1054, 2009.

[29] D. Nguyen, M. Dao, and T. Tran, "Video error concealment using sparse recovery and local dictionaries," in *Proceedings of ICASSP*, May 2011, pp. 1125–1128.

[30] Y. Zhang, X. Xiang, D. Zhao, S. Ma, and W. Gao, "Packet video error concealment with auto regressive model," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 12–27, January 2012.

[31] J. Koloda, J. Østergaard, S. Jensen, A. Peinado, and V. Sánchez, "Sequential error concealment for video/images via weighted template matching," in *Proceedings of DCC*, 2012, pp. 159–170.

[32] D. Giacobello, M. Christensen, M. Murthi, S. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE Transactions on Audio, Speech and Language Processing*, 2010.

[33] D. Donoho and Y. Tsaig, "Fast solution of L1-norm minimization problems when the solution may be sparse," *IEEE Transactions on Information Theory*, vol. 54, pp. 4789–4812, 2008.

[34] R. Little and D. Rubin, "Statistical analysis with missing data," Wiley, 1987.

[35] J. Romberg, "Imaging via compressive sensing," *IEEE Signal Processing Magazine*, vol. 25, no. 2, March 2008.

[36] L. Vandenberghe and S. Boyd, "Semidefinite programming," *Society for Industrial and Applied Mathematics*, 1996.

[37] X. Zhang, Y. Zhang, D. Zhao, S. Ma, and W. Gao, "A high efficient error concealment scheme based on auto-regressive model for video coding," in *PCS*, 2009.
[38] Z. Wang, E. Simoncelli, and A. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Signals, Systems and Computers*, vol. 2, pp. 1398–1402, November 2003.
[39] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assesment: From error visibility to structural visibility," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.
[40] J. Østergaard, M. Derpich, and S. Channappayya, "The high-resolution rate-distortion function under the structural similarity index," *EURASIP Journal on Advances in Signal Processing*, 2011.
[41] A. Brooks, X. Zhao, and T. Pappas, "Structural similarity quality metrics in a coding context: Exploring the space of realistic distorsions," *IEEE Transactions on Image Processing*, vol. 17, pp. 1261–1273, August 2008.
[42] V. Varsa and M. Hannuksela, "Non-normative error concealment algorithms," *ITU-T SG16, VCEG-N62*, vol. 50, September 2001.
[43] [Online], *Available: http://dtstc.ugr.es/~jkoloda/download.html*.

Ján Koloda received the M.Sc. degree in telecommunications engineering from the University of Granada, Granada, Spain, in 2009. He is currently working towards the Ph.D. degree on error concealment algorithms for block-coded video. Since 2010, he has been with the Research Group on Signal Processing, Multimedia Transmission and Speech/Audio Technologies, Department of Signal Theory, Networking and Communications at the University of Granada, under a research grant. He has been a visiting researcher at Aalborg University, Aalborg, Denmark. His research interests are in the area of error concealment of block-coded video sequences, image and signal processing.

Jan Østergaard (S'98-M'99-SM'11) received the M.Sc. degree in Electrical Engineering from Aalborg University, Aalborg, Denmark, in 1999 and the Ph.D. degree (cum laude) from Delft University of Technology, Delft, The Netherlands, in 2007. From 1999 to 2002, he worked as an R&D Engineer at ETI A/S, Aalborg, Denmark, and from 2002 to 2003, he worked as an R&D Engineer at ETI Inc., Virginia, United States. Between September 2007 and June 2008, he worked as a post-doctoral researcher at The University of Newcastle, NSW, Australia. From June 2008 to March 2011, he worked as a post-doctoral researcher/Assistant Professor at Aalborg University and he is currently Associate Professor at Aalborg University. He has also been a visiting researcher at Tel Aviv University, Tel Aviv, Israel, and at Universidad Técnica Federico Santa María, Valparaíso, Chile. He has received a Danish Independent Research Council's Young Researcher's Award and a post-doctoral fellowship from the Danish Research Council for Technology and Production Sciences.

Søren Holdt Jensen (S'87-M'88-SM'00) received the M.Sc. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1988, and the Ph.D. degree in signal processing from the Technical University of Denmark, Lyngby, Denmark, in 1995. Before joining the Department of Electronic Systems of Aalborg University, he was with the Telecommunications Laboratory of Telecom Denmark, Ltd, Copenhagen, Denmark; the Electronics Institute of the Technical University of Denmark; the Scientific Computing Group of Danish Computing Center for Research and Education, Lyngby; the Electrical Engineering Department of Katholieke Universiteit Leuven, Leuven, Belgium; and the Center for PersonKommunikation (CPK) of Aalborg University. He is Full Professor and is currently heading a research team working in the area of numerical algorithms, optimization, and signal processing for speech and audio processing, image and video processing, multimedia technologies, and digital communications. Prof. Jensen was an Associate Editor for the IEEE Transactions on Signal Processing and Elsevier Signal Processing, and is currently Associate Editor for the IEEE Transactions on Audio, Speech and Language Processing. He is a recipient of an European Community Marie Curie Fellowship, former Chairman of the IEEE Denmark Section and the IEEE Denmark Sections Signal Processing Chapter. He is member of the Danish Academy of Technical Sciences and was in January 2011 appointed as member of the Danish Council for Independent Research—Technology and Production Sciences by the Danish Minister for Science, Technology and Innovation.

Victoria Sánchez (M'95) received the M.S. and the Ph.D. degrees from the University of Granada, Granada, Spain, in 1988 and 1995, respectively. In 1988, she joined the Signal Processing and Communications department of the University of Granada where she is currently a member of the research group on Signal Processing, Multimedia Transmission and Speech/Audio Technologies. During 1991, she was visiting with the Electrical Engineering Department, University of Sherbrooke, Canada. Since 1997, she is an Associate Professor at the University of Granada. Her research interests include speech and audio processing, multimedia transmission and speech recognition. She has authored over 60 journal articles and conference papers in these fields.

Antonio M. Peinado (M'95-SM'05) received the M.S. and the Ph.D. degrees in Physics from the University of Granada, Granada, Spain, in 1987 and 1994, respectively. Since 1988, he has been working at the University of Granada, where he has led several research projects related to signal processing and communications. In 1989, he was a Consultant at the Speech Research Department, AT&T Bell Labs. He earned the positions of Associate Professor (1996) and Full Professor (2010) in the Department of Signal Theory, Networking and Communications, University of Granada, and is currently director of the research group on Signal Processing, Multimedia Transmission and Speech/Audio Processing (SigMAT) at the same university. He is the author of numerous publications and coauthor of the book Speech Recognition over Digital Channels (Wiley, 2006), and has served as reviewer for international journals, conferences and project proposals. His current research interests are focused on robust speech recognition and transmission, robust video transmission, and ultrasound signal processing.