ON THE APPLICATION OF MULTIVARIATE KERNEL DENSITY ESTIMATION TO IMAGE ERROR CONCEALMENT

Ján Koloda, Antonio M. Peinado, Victoria Sánchez

Dpt. Teoría de la Señal, Telemática y Comunicaciones - CITIC
Universidad de Granada, Spain
{janko,amp,victoria}@ugr.es

ABSTRACT

This paper proposes a methodology for the application of multivariate kernel density estimation (KDE) to MMSE-based image/video error concealment (EC). We show that the estimation of the kernel bandwidth matrix for EC must follow a criterion different from that of typical KDE problems. In particular, we propose a bandwidth built as the product of a structure matrix and a scale factor obtained with a minimum square error criterion. We show that our proposal can achieve average PSNR improvements larger than 1 dB with respect to other state-of-the-art techniques.

***Index Terms*—** kernel estimation, error concealment

1. INTRODUCTION

Achieving high QoS in multimedia applications is a very challenging task since the transmission of multimedia contents over error prone channels may lead to errors or data losses. The most advanced and utilized image and video coding systems (JPEG, H.264/AVC, etc.) are block-based so these errors result in a loss of one or several macroblocks. In order to mitigate the effect of these losses, error concealment (EC) algorithms can be applied at the decoder. They take advantage of spatial and/or temporal correlations within the received stream to recover the missing data. For image communication or video transmission, when temporal information is not available or relevant, only spatial EC (SEC) is applicable.

A simple and common SEC technique is bilinear interpolation [1] which is defined as the default SEC method in the H.264/AVC codec. In order to better preserve important visual features, such as edges, a more advanced technique based on Markov random fields was proposed in [2]. In [3], a sequential pixel-wise method that draws on orientation adaptive interpolation was introduced. Bilateral filtering that exploits a pair of gaussian kernels is treated in [4]. A switching content adaptive SEC algorithm was proposed in [5]. Inpainting methods have also been successfully applied to EC problems [6]. A Hough transform based technique that aims at recovering edges based on their visual properties was proposed in [7]. Also, SEC techniques in a transformed domain have been recently proven to produce high-quality reconstructions [8].

In our previous paper [9] we proposed an EC technique which estimates a lost group of pixels (patch) through linear prediction (LP). This method provided better results than other state-of-the-art techniques such as [1]-[8]. The LP predictor is obtained by minimizing the square error between a context vector containing the available pixels around the missing patch and a linear combination of context vectors taken from the neighbourhood. This optimization was carried out under constraints of non-negativity and sparsity via convex relaxation. We also showed that the resulting estimation could be approximated by a multivariate Nadaraya-Watson regression with a Gaussian kernel [10]. This kernel-based view of sparse linear prediction offers a number of advantages. In particular, it can be interpreted as a minimum mean square error (MMSE) estimation where the required probabilities have been obtained through kernel density estimation (KDE) [11]. In this paper, we will exploit and generalize this new point of view which will allow us to apply powerful Bayesian tools to EC. Moreover, we will see that the goal of signal reconstruction is quite different from that of regression. Since the main problem in KDE is the estimation of the bandwidth matrix H, this means that H must be computed with a criterion different from the one usually applied for KDE or regression. Thus, we will propose a method to obtain the bandwidth which is specifically conceived for reconstruction.

The paper is organized as follows. The EC framework is detailed in Section 2. The proposed algorithm is described in Sections 3 and 4. Simulations results are discussed in Section 5. The last section is devoted to conclusions.

2. PREVIOUS WORK AND CONCEALMENT FRAMEWORK

The concealment framework used along this paper will be the same as that of reference [9]. In the following, we briefly summarize it. Let $\mathcal{L}$ be the set of missing pixels. Our goal

This work has been supported by an FPU grant from the Spanish Ministry of Education and by the MICINN TEC2010-18009 project.

is the prediction of a vector $\mathbf{z}_0 = (\mathbf{x}_0^t, \mathbf{y}_0^t)^t$, where $\mathbf{x}_0$ is a patch of lost pixels in $\mathcal{L}$ and $\mathbf{y}_0$ contains a set of (adjacent and available) context pixels. Let $\mathcal{S}$ be the set of available pixels which can be employed for prediction. We will consider all the possible vectors $\mathbf{z}_j$ ($j = 1, \ldots, M$) that can be built in $\mathcal{S}$ with the same shape and dimensionality as $\mathbf{z}_0$ (that is, $\mathbf{z}_j = (\mathbf{x}_j^t, \mathbf{y}_j^t)^t$). Then, the LP estimator for $\mathbf{x}_0$ can be written as,

$$\hat{\mathbf{x}}_0 = \sum_{j=1}^{M} w_j \mathbf{x}_j, \tag{1}$$

where $\mathbf{w} = (w_1, \ldots, w_M)^t$ is the vector of LP coefficients.

We consider a block-based codec where the missing region $\mathcal{L}$ is a 16×16 macroblock and the support area $\mathcal{S}$ comprises all the available pixels within the neighbouring macroblocks around $\mathcal{L}$. In this paper, we will employ an error pattern as shown in Fig. 3(a) which corresponds to a rate of block loss of approximately 25% with dispersed slicing structure [12]. Note, however, that our technique can be straightforwardly extended to other error patterns. We will also consider 2×2 patches $\mathbf{x}_0$ of missing pixels and the corresponding context $\mathbf{y}_0$ will comprise all the available pixels within the 6×6 pixel neighbourhood centred in $\mathbf{x}_0$. Vectors $\mathbf{z}_j$ replicate the shape of $\mathbf{z}_0$. These configurations are shown in Fig. 1(a). Moreover, macroblocks are concealed sequentially from the outer layer towards the centre (see Fig.1(b)). This filling order is based on a reliability parameter and it is detailed in [9].

In our previous work [9], the weights w_j of Eq. (1) are obtained by minimizing the following square error,

$$\epsilon_{\mathbf{y}}(\mathbf{w}; \mathbf{y}_0) = \left\| \mathbf{y}_0 - \sum_{j=1}^{M} w_j \mathbf{y}_j \right\|_2^2 \tag{2}$$

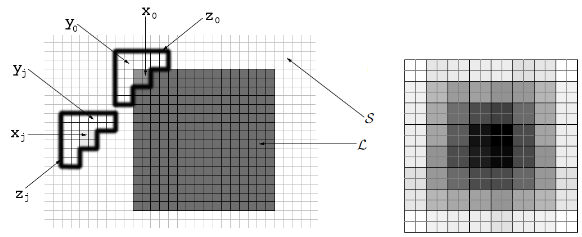
along with non-negativity ($\mathbf{w} \succeq 0$) and sparsity constraints. In [9], we also showed that these LP weights could be approximated through the following exponential function,

$$w_j = C \exp\left(-\frac{1}{2} \frac{\|\mathbf{y}_0 - \mathbf{y}_j\|^2}{m\sigma^2} \right), \tag{3}$$

where σ^2 is a decay factor ($\sigma^2 = 10$ in [9]), m is the dimensionality of the context vectors, and C is a normalization factor so that $\sum_j w_j = 1$. The resulting estimation can be viewed as a particular form of Nadaraya-Watson regression which employs a multivariate Gaussian kernel with a scalar bandwidth $h = \sqrt{m\sigma^2}$. This new kernel-based point of view is exploited in the following section.

3. THE KERNEL-BASED APPROACH

Kernel density estimation (KDE) is a non-parametric way for the estimation of the probability density function (pdf) associated to a given random process from a set of observations. In our case, we are interested in the pdf of $\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)^t$ from the set of observations $\{\mathbf{z}_j; j = 1, \ldots, M\}$. The corresponding



(a) (b)

Fig. 1. (a) Example of configuration for the vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$. $\mathcal{S}$ 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. The regions illustrated by brighter level are recovered first.

KDE estimate can be written as,

$$p(\mathbf{z}) = \frac{1}{M} \sum_{j=1}^{M} \frac{1}{|H|} K\left(H^{-1}(\mathbf{z} - \mathbf{z}_j)\right) = \frac{1}{M} \sum_{j=1}^{M} K_Z^{(j)}(\mathbf{z}). \tag{4}$$

where $K(\mathbf{u}) = \exp(-\mathbf{u}^t\mathbf{u}/2)/\sqrt{2\pi}$ is the (Gaussian) kernel employed and H is the bandwidth matrix. A more convenient form of the KDE estimator is given in the last part of Eq. (4), where $p(\mathbf{z})$ adopts the form of a Gaussian mixture model (GMM) and $K_Z^{(j)}(\mathbf{z})$ represents a multivariate Gaussian with mean $\mathbf{z}_j$ and covariance $\mathcal{H}$ which can be decomposed as,

$$\mathcal{H} = HH^t = \begin{pmatrix} \mathcal{H}_{XX} & \mathcal{H}_{XY} \\ \mathcal{H}_{YX} & \mathcal{H}_{YY} \end{pmatrix}. \tag{5}$$

In the following, we will also refer to $\mathcal{H}$ as bandwidth matrix.

Once $p(\mathbf{z})$ has been obtained, different Bayesian estimation techniques can be carried out. In particular, we are interested in the MMSE estimator of $\mathbf{x}_0$ given $\mathbf{y}_0$. Since the KDE estimate has the form of a GMM, we can adapt the well-known MMSE estimation formulae for GMM models [13], obtaining

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}|\mathbf{y}_0] = \textstyle\sum_{j=1}^{M} w_j(\mathbf{y}_0) \boldsymbol{\mu}_{X|Y}^{(j)}(\mathbf{y}_0) \tag{6}$$

$$w_j(\mathbf{y}_0) = \frac{K_Y^{(j)}(\mathbf{y}_0)}{\sum_{i=1}^{M} K_Y^{(i)}(\mathbf{y}_0)} \tag{7}$$

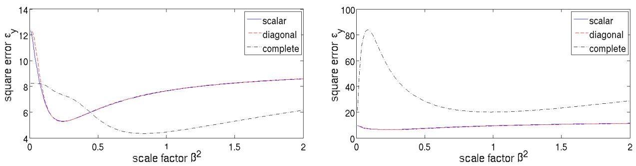
$$\boldsymbol{\mu}_{X|Y}^{(j)}(\mathbf{y}_0) = E[\mathbf{x}|\mathbf{y}_0, \mathbf{y}_j] = \mathbf{x}_j + \mathcal{H}_{XY}\mathcal{H}_{YY}^{-1}(\mathbf{y}_0 - \mathbf{y}_j) \tag{8}$$

where $K_Y^{(j)}(\mathbf{y})$ represents a multivariate Gaussian with mean $\mathbf{y}_j$ and covariance $\mathcal{H}_{YY}$. The estimator just derived can be interpreted as a multivariate generalization of the Nadaraya-Watson regressor defined by Eqs. (1) and (3). Note that, unlike the MMSE estimator in [13], our proposal does not require an off-line GMM training and can be easily applied on-line from the set of available vectors $\mathbf{z}_j (j = 1, ..., M)$.

4. BANDWIDTH ESTIMATION

4.1. Classical KDE estimation

The most important issue in KDE problems is the bandwidth estimation (BE). There exist several approaches for it. A pop-



(a) (b)

Fig. 2. Example of $\epsilon_{\mathbf{y}} = \epsilon_{\mathbf{y}}(\beta^2; \mathbf{y}_0)$ for two different patches using scalar, diagonal and complete bandwidths.

ular and usually recommended approach is that of the so-called plug-in methods [14]. The goal of these methods is the minimization of the asymptotic mean integrated squared error (AMISE).

In this paper, we will consider the plug-in method for multivariate KDE described in [15, 16]. In this case, it is considered that $\mathcal{H} = \beta^2 F_{ZZ}$, where β^2 is a scale factor and F_{ZZ} is a structure matrix. If F_{ZZ} is known, the problem is reduced to the estimation of the scale factor β^2. This requires a quite complex procedure whose details can be found in [15]. It is interesting to note that if we decompose F_{ZZ} in the same way as in (5) the conditional mean of Eq.(8) does not depend on β^2, that is,

$$\boldsymbol{\mu}_{X|Y}^{(j)}(\mathbf{y}_0) = E[\mathbf{x}|\mathbf{y}_0, \mathbf{y}_j] = \mathbf{x}_j + F_{XY} F_{YY}^{-1} (\mathbf{y}_0 - \mathbf{y}_j), \quad (9)$$

In [15], F_{ZZ} is approximated by the covariance matrix C_{ZZ} of the observed samples $\{\mathbf{z}_j; j = 1, \ldots, M\}$.

4.2. A minimum square error (MSE) approach

The classical methods for BE in KDE (or regression) problems try to estimate a pdf suitable for the whole space of observations [14]. However, the goal of reconstruction techniques is to obtain an estimate of a specific patch $\mathbf{x}_0$ given its known context $\mathbf{y}_0$. Thus, a BE procedure to be employed in signal recovery problems should be oriented to be as accurate as possible at the point of interest.

In this paper we propose that the criterion for BE should be the same as the one employed for the sparse linear prediction method in [9], that is, the minimization of the square error of Eq. (2). This minimization is now constrained to weights of the form given by Eq. (7). Since these weights only depend on the bandwidth $\mathcal{H}$, we can consider that the function to be minimized is $\epsilon_{\mathbf{y}} = \epsilon_{\mathbf{y}}(\mathcal{H}; \mathbf{y}_0)$.

In order to carry the minimization of $\epsilon_{\mathbf{y}}$ versus the bandwidth $\mathcal{H}$ we could apply some sort of optimization algorithm. Some preliminary experiments (with a steepest descent procedure) have revealed that this type of solution yields an unstable convergence and poor results due to the large number of parameters in matrix $\mathcal{H}$. Only in the case of considering a scalar bandwidth (that is, $\mathcal{H} = h^2 I$, I identity matrix), we could obtain acceptable results. However, even in this case, the steepest descent solution was not worthwhile either since the minimization of $\epsilon_{\mathbf{y}} = \epsilon_{\mathbf{y}}(h^2; \mathbf{y}_0)$ was even much more time-consuming than an exhaustive search within the typical range of variation of h^2.

In order to overcome these problems, in this paper we propose a BE procedure as follows:

1. We will adopt the same assumption as in the plug-in BE method described above based on the use of a scale factor β^2 and a known structure F_{ZZ}, that is, $\mathcal{H} = \beta^2 F_{ZZ}$.
2. Then, since F_{ZZ} is fixed, the weights are only functions of β^2 (that is, $w_j = w_j(\beta^2; \mathbf{y}_0)$), so that the square error to be minimized $\epsilon_{\mathbf{y}} = \epsilon_{\mathbf{y}}(\beta^2; \mathbf{y}_0)$ also depends only on the scale factor β^2. Therefore, the corresponding minimization is feasible by exhaustive search within the typical range of variation of β^2.

In order to carry out an efficient exhaustive search, we can define a set of auxiliary weights as follows,

$$\tilde{w}_j(\mathbf{y}_0) = \exp\left((\mathbf{y}_0 - \mathbf{y}_j)^t F_{YY}^{-1} (\mathbf{y}_0 - \mathbf{y}_j)\right). \quad (10)$$

These auxiliary weights do not depend on β^2 and can be precomputed. Then, during the exhaustive search, the weights (Eq.(7)) for every value of β^2 can be efficiently obtained as,

$$w_j(\beta^2; \mathbf{y}_0) = \frac{(\tilde{w}_j(\mathbf{y}_0))^{1/\beta^2}}{\sum_{i=1}^{M} (\tilde{w}_i(\mathbf{y}_0))^{1/\beta^2}}. \quad (11)$$

Finally, once the optimal value of β^2 and its corresponding weights have been obtained, the unknown patch $\mathbf{x}_0$ can be estimated through Eqs. (6) and (9).

Several approaches for BE are adopted depending on the selection of the structure matrix F_{ZZ} according to its level of complexity [17]:

1. A scalar bandwidth $F_{ZZ} = \sigma_Z^2 I$, where σ_Z^2 is the variance of the available pixels (in set $\mathcal{S}$). This approach can be reduced to the algorithm described in [9] by forcing $\beta^2 \sigma_Z^2 = 10m$.
2. A diagonal bandwidth $F_{ZZ} = diag(C_{ZZ})I$.
3. A complete bandwidth $F_{ZZ} = C_{ZZ}$, as in [15].

Figure 2(a) shows examples of the error curve $\epsilon_{\mathbf{y}}(\beta^2; \mathbf{y}_0)$, obtained during the minimization procedure, for all three approaches. Scalar and diagonal bandwidths produce almost identical results since $\boldsymbol{\mu}_{X|Y}^{(j)}(\mathbf{y}_0) = \mathbf{x}_j$ for both cases and the diagonal of the correlation matrix C_{ZZ} tends to be uniform. Simulations reveal that both configurations tend to smooth high frequency textures (see Fig. 3(b)). On the other hand, complete bandwidth matrices can recover fine textures with high accuracy (see Fig.3(c)), although sometimes they show an unexpected behaviour (see Fig. 2(b)). A possible explanation is that the minimization of $\epsilon_{\mathbf{y}}(\beta^2; \mathbf{y}_0)$ is equivalent to the maximization of the corresponding PSNR only if $F_{YY} = \sigma_Z^2 I$. Moreover, the scalar (and diagonal) approach is more robust against non-stationarity. In this case, an inaccurate selection of the structure matrix $F_{ZZ} = \sigma_Z^2 I$ can be corrected by modifying β^2, since $\mathcal{H} = \beta^2 \sigma_Z^2 I$. This, however, is not possible for complete structure matrices. Thus, in order to achieve a compromise between texture reconstruction and PSNR, we will also test a combination of scalar and

SEC	*Lena*		*Goldhill*		*Barbara*		Average	
	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM	PSNR	MS-SSIM
[1]	30.42	96.56	31.27	95.65	26.85	94.87	28.57	95.04
[5]	31.96	97.25	30.24	94.53	27.39	96.20	29.46	95.71
[2]	32.17	97.64	31.12	95.71	27.99	96.00	29.57	95.99
[4]	32.15	97.44	30.91	95.52	29.91	97.04	30.22	96.39
[6]	30.85	97.08	30.40	95.21	28.03	95.72	29.23	95.69
[7]	32.70	97.96	31.66	96.35	28.41	97.37	30.28	96.74
[3]	32.82	97.65	31.54	95.62	29.66	97.07	30.35	96.08
[8]	32.72	97.80	31.78	96.14	30.84	97.64	30.50	96.56
[9]	32.55	97.97	31.72	96.43	30.80	98.01	30.55	96.98
KD_S	32.22	98.02	31.43	96.40	30.84	98.11	30.51	97.00
MS_S	32.84	98.11	32.03	96.67	31.33	98.25	31.20	97.27
MS_D	32.87	98.11	32.02	96.66	31.35	98.26	31.21	97.28
MS_C	32.69	98.00	32.14	96.77	31.77	98.35	31.24	97.28
MS_X	**33.00**	**98.18**	**32.17**	**96.84**	**32.22**	**98.55**	**31.43**	**97.40**

Table 1. 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.

complete bandwidths where the complete bandwidth matrix is employed to compute the conditional means $\boldsymbol{\mu}_{X|Y}^{(j)}(\mathbf{y}_0)$ (Eq. (8)) and the scalar bandwidth for the weights w_j (Eq. (7)).

5. EXPERIMENTAL RESULTS

In order to reflect the perceptual quality of the reconstructions, the multi-scale structural similarity (MS-SSIM) index [18] is used for comparison along with the objective PSNR measure. MS-SSIM is a weighted combination of SSIM indices computed over different image resolutions. Thus, coarse structures as well as fine textures are taken into account. SSIM index aims at approximating the human visual system response looking for similarities in structure, contrast and intensity [19].

The performance of our different proposals is tested on the images of *Lena* (512×512), *Goldhill* (720×576), *Foreman* (352×288), *Barbara* (512×512), *Baboon* (512×512), *Clown* (512×512), *Tire* (205×232), *Pirate* (1024×1024), *Boat* (512×512) and *Peppers* (384×512). We will use the framework described in Section 2. For our MSE approaches, β^2 is searched exhaustively within the range [0, 2] with steps of 0.01. First, we test the performance of the scalar bandwidth using the classical KDE (KD_S) of Section 4.1 and our MSE approach (MS_S) and compare them with our previous exponential sparse linear prediction (SLP) algorithm [9], which we will use as a reference (marked in Table 1). Table 1 shows that KD_S performs considerably worse than MS_S and it provides virtually no improvement over [9]. Thus, in the following, we focus on the MSE approach for scalar bandwidth as well as diagonal (MS_D) and complete (MS_C) bandwidth matrices. Moreover, we also use the combined scenario with scalar and complete bandwidths (MS_X) as described at the end of Section 4.2. We compare our proposals with other state-of-the-art SEC techniques [1]-[9].

Table 1 shows the results in terms of PSNR and MS-SSIM for the images of *Lena*, *Goldhill* and *Barbara* as well as the

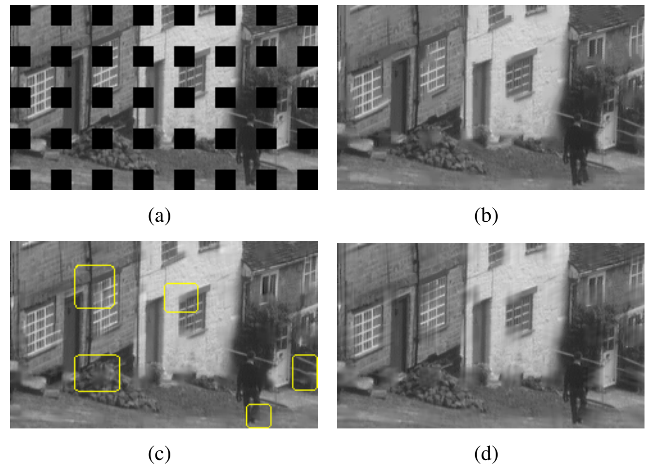


Fig. 3. Subjective comparison for a fraction of *Goldhill*. (a) Received data. (b) Reconstruction using scalar bandwidth (MS_S). (c) Reconstruction using complete bandwidth matrix (MS_C). (d) Reconstructed by [8].

average performance over all ten tested images. The results confirm our hypothesis that the KD_S approach is not EC oriented. On the other hand, all of our MSE proposals outperform the other techniques, including our previous exponential SLP. In addition, scalar and diagonal bandwidths produce almost identical results. The complete bandwidth performs better on average although is inferior in some particular cases (e.g. *Lena*). Finally, the combination of the high quality reconstructions produced by complete bandwidth with the good behaviour of the scalar one produces the best result both on subjective and objective levels. A subjective comparison is shown in Fig.3.

Due to the efficient implementation of the exhaustive search, carried out utilizing precomputed weights (Eqs. (10) and (11)), the computational complexity is only moderately increased with respect to SLP for all the MSE proposals. This increment of complexity is reflected in the reconstruction quality which is improved in 0.9dB on average (for MS_X). The KD_S approach, on the other hand, requires up to half an hour per macroblock and therefore is computationally prohibitive for on-line applications.

6. CONCLUSIONS

We have proposed a framework for image EC based on a generalization of the Nadaraya-Watson estimator with an MSE-based bandwidth estimation. We have shown that this MSE criterion achieves a performance significantly better than that of classical KDE with pdf matching. Using a simple scalar bandwidth we achieve an average improvement over [9] of 0.7dB. This improvement is later incremented up to almost 1dB by combining the robustness of the scalar bandwidth with the accurate reconstructions of fine textures produced by complete bandwidth matrices. Ongoing work is focused on a more accurate selection of the bandwidth matrix structure.

7. REFERENCES

[1] T. Wiegand, G.J. 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.

[2] 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*, 1999, vol. 6, pp. 3117–3120.

[3] X. Li and M.T. 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.

[4] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Image error-concealment via block-based bilateral filtering," *IEEE International Conference on Multimedia and Expo*, pp. 621–624, June 2008.

[5] 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.

[6] P.F. Harrison, *Texture Synthesis, Texture Transfer and Plausible Restoration*, Ph.D. thesis, Monash University, 2005.

[7] J. Koloda, V. Sánchez, and A. M. Peinado, "Spatial error concealment based on edge visual clearness for image/video communication," *Circuits, Systems and Signal Processing*, October 2012.

[8] A. Kaup, K. Meisinger, and T. Aach, "Frequency selective signal extrapolation with applications to error concealment in image communication," *AEUE - International Journal of Electronics and Communications*, vol. 59, pp. 147–156, 2005.

[9] J. Koloda, J. Østergaard, S. H. Jensen, V. Sánchez, and A. M. Peinado, "Sequential error concealment for video/images by sparse linear prediction," *IEEE Transactions on Multimedia*, In Press.

[10] E.A. Nadaraya, "On estimating regression," *Theory of Probability and its Applications*, vol. 9, pp. 141–142, September 1964.

[11] D.W. Scott, "Multivariate density estimation: Theory, practice, and visualization," Wiley, 1992.

[12] ITU-T, "ITU-T Recommendation H.264," International Telecommunication Union, 2010.

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

[14] S.J. Sheather, "Density estimation," *Statistical Science*, vol. 19, no. 4, pp. 588–597, 2004.

[15] M. Kristan, A. Leonardis, and D. Skočaj, "Multivariate online kernel density estimation with gaussian kernels," *Pattern Recognition*, vol. 44, pp. 2630–2642, 2011.

[16] M.P. Wand and M.C. Jones, "Multivariate plug-in bandwidth selection," *Computational Statistics*, , no. 9, pp. 97–117, 1994.

[17] M.P. Wand and M.C. Jones, "Comparison of smoothing parameterizations in bivariate kernel density estimation," *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 520–528, 1993.

[18] Z. Wang, E.P. Simoncelli, and A.C. Bovik, "Multiscale structural similarity for image quality assessment," *IEEE Signals, Systems and Computers*, vol. 2, pp. 1398–1402, November 2003.

[19] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assesment: From error visibility to structural visibility," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.