MULTIVARIATE STATISTICAL APPROACH FOR IMAGE DENOISING

Dongwook Cho and T. D. Bui

Concordia University
Department of Computer Science
1455 De Maisonneuve Blvd. W., Montréal, Québec H3G 1M8, Canada

ABSTRACT

In this paper, we derive the general estimation rule in the wavelet domain to obtain the denoised coefficients from the noisy image based on the multivariate statistical theory. We define a parametric multivariate generalized Gaussian distribution (MGGD) model which closely fits the actual distribution of wavelet coefficients in clean natural images. Multivariate model makes it possible to exploit the dependency between the estimated wavelet coefficients and their neighbours or other coefficients in different subbands. Also it can be shown that some of the existing methods based on statistical modeling are subsets of our multivariate approach. Our method could achieve high quality image denoising. Among the comparable image denoising methods using the same type of wavelet (esp. Daubechies 8) filter, our results produce comparatively higher peak signal to noise ratio (PSNR).

1. INTRODUCTION

Various recent works on image denoising using wavelet transforms have shown that wavelet is an efficient tool for noise removal of noisy images [1, 2, 3, 4, 5, 6]. Since wavelet transform divides an image domain into the low and high frequency domains by its filters, it is natural that some noise which is closely related to the high frequency domain can be removed by killing some small coefficients in the high frequency part.

Wavelet coefficients are not strongly correlated, but they still have dependency on each other. So many of the recent works such as [2, 3, 6, 7, 8] take into account their dependency in order to obtain a proper estimate. They show that incorporating different information like neighbours and parents is helpful to preserve details and remove noise for natural image denoising. In addition, we may consider some other information like coefficients in the finer decomposition level (*offsprings*) or corresponding coefficients of the other subbands in the same level. Hence, it is possible that we can estimate a denoised wavelet coefficient using all the related coefficients.

Our concerns lie in finding a general way to estimate the denoised wavelet coefficients using related coefficients such as neighbours, parents and offsprings. We use multivariate Bayesian approach, which is classical yet powerful for estimating the denoised coefficients. It can be shown that some of the existing methods based on statistical modeling are subsets of our multivariate approach by changing the related elements and varying the distribution parameters.

This work was supported by research grants from the Natural Sciences and Engineering Research Council of Canada and by the Fonds pour la Formation de Chercheurs et l'Aide à la Recherche of Quebec.

2. BAYESIAN ESTIMATION FOR MGGD MODEL

Let A be a clean natural image with size $N \times N$, B a noisy image which can be expressed as $B = A + \sigma C$, and C zero-mean Gaussian white noise, which is $C \sim N(0, 1)$. σ^2 is noise variance.

After performing multiresolution wavelet decomposition on B, we get the wavelet coefficient $y_{j,k}$, which is the k-th wavelet coefficient in j-th level for B. Due to the linearity of the wavelet transform, we have $y_{j,k} = x_{j,k} + \sigma z_{j,k}$, where $x_{j,k}$ and $z_{j,k}$ are the wavelet coefficients of A and C respectively in the same location as $y_{j,k}$.

Let $\mathbf{x}$ be a d-dimensional wavelet coefficient vector, $\mathbf{x} = (x_1, x_2, \cdots, x_d)^t$, where x_1 is the wavelet coefficient under consideration and x_i $(i = 2, \cdots, d)$ are the related coefficients to be taken into consideration, e.g. neighbours, parent and offsprings. Here for simplicity, we replace the double subscripts in $x_{j,k}$ by a single subscript x_i. The corresponding vectors $\mathbf{y}$ and $\mathbf{z}$ can be similarly defined for the noisy image B and the noise C. We assume that x_i, y_i and z_i correspond to each other in both decomposition level and location. Hence we have $\mathbf{y} = \mathbf{x} + \sigma\mathbf{z}$. For the sake of simplicity, we omit subscripts j, k in the rest of the paper.

Our concern lies mainly in estimating the unknown wavelet coefficient vector $\hat{\mathbf{x}}$, and $\hat{\mathbf{x}}$ should be obtained only from $\mathbf{y}$ of the noisy image B. One of the ways to estimate $\hat{\mathbf{x}}$ is to use MAP estimator to maximize $p(\mathbf{x}|\mathbf{y})$.

Theorem. *By MAP estimator, the estimated clean vector $\hat{\mathbf{x}}$ can be expressed as :*

$$\hat{\mathbf{x}} = \mathbf{y} + \sigma^2 \nabla g(\hat{\mathbf{x}}), \tag{1}$$

where $g(\hat{\mathbf{x}})$ is $\ln p(\hat{\mathbf{x}})$.

Proof. From the usual MAP estimator for $\hat{\mathbf{x}}$ can be obtained as follows:

$$\begin{aligned} \hat{\mathbf{x}} &= \arg\max_{\mathbf{x}\in\mathbb{R}^d} \ln p(\mathbf{x}|\mathbf{y}) \\ &= \arg\max_{\mathbf{x}\in\mathbb{R}^d} \left[\ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})\right]. \end{aligned} \tag{2}$$

Eq. (2) means that the optimal value $\hat{\mathbf{x}}$ with minimum probability error can be estimated by $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$. $p(\mathbf{y}|\mathbf{x})$ is the multivariate Gaussian distribution with $N(\mathbf{0}, \Sigma_z = \sigma^2\mathbf{I})$ since Gaussian noise is independently and identically distributed for each element of the vector. Hence,

$$\ln p(\mathbf{y}|\mathbf{x}) = -\frac{d}{2}\ln\left(2\pi\sigma^2\right) - \frac{(\mathbf{y}-\mathbf{x})^t(\mathbf{y}-\mathbf{x})}{2\sigma^2} \tag{3}$$

We assume that $p(\mathbf{x})$ is known. $p(\mathbf{x})$ might vary depending on the type of sample images. From Eqs. (2) and (3) :

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}\in\mathbb{R}^d}\left[-\frac{d}{2}\ln\left(2\pi\sigma^2\right)-\frac{(\mathbf{y}-\mathbf{x})^t(\mathbf{y}-\mathbf{x})}{2\sigma^2}+g(\mathbf{x})\right]\\ &= \arg\max_{\mathbf{x}\in\mathbb{R}^d} F(\mathbf{x}), \qquad (4)\end{aligned}$$

where $g(\mathbf{x}) = \ln p(\mathbf{x})$ and $F(\mathbf{x})$ represents the term inside $\arg\max$. If there exists $\hat{\mathbf{x}}$ which satisfies $F(\hat{\mathbf{x}}) > \lim_{x_i\to\pm\infty} F(\mathbf{x})$, Eq. (4) is equivalent to the solution of the following equation:

$$\begin{aligned}\nabla F(\hat{\mathbf{x}}) = -\frac{\hat{\mathbf{x}}-\mathbf{y}}{\sigma^2}+\nabla g(\hat{\mathbf{x}}) = 0\\ \Leftrightarrow \hat{\mathbf{x}} = \mathbf{y}+\sigma^2\nabla g(\hat{\mathbf{x}}). \qquad (5)\end{aligned}$$

This completes the proof. □

The existing models for wavelet denoising are usually based on univariate statistical model whereas $p(\mathbf{x})$ is a multivariate pdf in our model. There are several multivariate functions which are symmetric spherically like multivariate Gaussian model. In our paper, we use extended generalized Gaussian distribution (GGD) model [9] for its simple form and to achieve good fitting errors. We call this model multivariate generalized Gaussian distribution (MGGD):

$$p(\mathbf{x}) = \gamma \exp\left\{-\left(\frac{(\mathbf{x}-\mu)^t\Sigma_{\mathbf{x}}^{-1}(\mathbf{x}-\mu)}{\alpha}\right)^{\beta}\right\}, \qquad (6)$$

where α and β are parameters which can represent the spherical shape of the model and γ indicates a normalized constant defined by α, β and the covariance matrix $\Sigma_{\mathbf{x}}$.

When the dimension of $\mathbf{x}$ is one (scalar), the MGGD is still applicable and is denoted by UGGD. MGGD is a particular case of the v-spherical distribution defined by Fernández [10]. Using MGGD model, we can derive more specific forms of Eq. (1). Since we can assume that $\mu = \mathbf{0}$,

$$\nabla g(\mathbf{x}) = -\frac{2\beta}{\alpha^\beta}(\mathbf{x}^t\Sigma_{\mathbf{x}}^{-1}\mathbf{x})^{\beta-1}\Sigma_{\mathbf{x}}^{-1}\mathbf{x}. \qquad (7)$$

From Eqs. (1) and (7),

$$\hat{\mathbf{x}} = \left(\Sigma_{\hat{\mathbf{x}}}+\frac{2\sigma^2\beta}{\alpha^\beta}(\hat{\mathbf{x}}^t\Sigma_{\hat{\mathbf{x}}}^{-1}\hat{\mathbf{x}})^{\beta-1}I\right)^{-1}\Sigma_{\hat{\mathbf{x}}}\mathbf{y}. \qquad (8)$$

To simplify Eq. (8), we define $q(\hat{\mathbf{x}}) = \hat{\mathbf{x}}^t\Sigma_{\hat{\mathbf{x}}}^{-1}\hat{\mathbf{x}}$. Hence :

$$q(\hat{\mathbf{x}}) = \mathbf{y}^t\left(\Sigma_{\hat{\mathbf{x}}}+\frac{2\sigma^2\beta\{q(\hat{\mathbf{x}})\}^{\beta-1}}{\alpha^\beta}I\right)^{-2}\Sigma_{\hat{\mathbf{x}}}\mathbf{y}. \qquad (9)$$

Eqs. (8) and (9) allow us to solve for $\hat{\mathbf{x}}$.

However, there is no general solution for Eq. (9). To overcome this problem, we can define a particular condition for α, β and $\Sigma_{\hat{\mathbf{x}}}$ or use a numerical method.

3. MODEL SELECTION

Although we have derived the general formula for $\hat{\mathbf{x}}$, the best fit parameters α and β for MGGD model in Eq. (6) should be decided. The parameters could be estimated by analyzing the distribution of the wavelet coefficients for natural images. It is well-known that the distribution of detailed wavelet coefficients from natural images looks Gaussian-like with zero mean such as GGD [9]. We have tried to find the closest MGGD model for each subband analytically. 20 test images with 512×512 size have been inspected to extract enough sample coefficients. The test images are from free images collected and offered by Computer Vision Group in University of Granada, Spain: `http://decsai.ugr.es/cvg`. In this analysis, Daubechies 8 filter is applied to the image set for the wavelet decomposition. The distribution parameters could be slightly different depending on the mother wavelet.

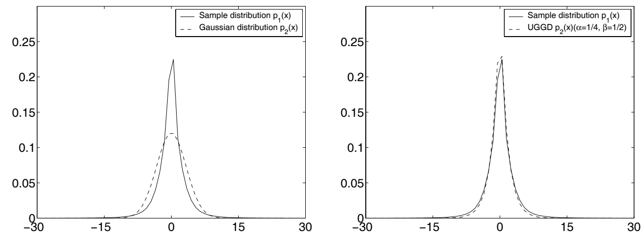

Fig. 1. Distribution of sample coefficients and the estimated UGGD function (using sample coefficients in HH subband of the 1st decomposition level).

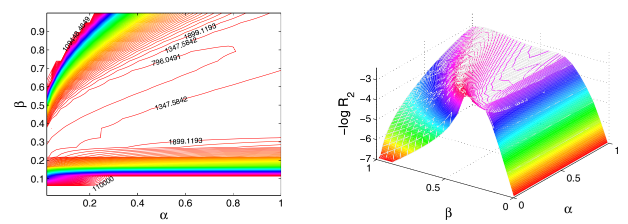

Fig. 2. Norm of residue R contour (left) and $-\log R$ projection (right) for varying MGGD parameters α and β for bivariate model.

Fig. 1 shows the difference between the sample distribution and its estimated pdfs. As mentioned before, Gaussian distribution in the left figure does not fit the sample distribution closely but UGGD model with particular parameters is better adapted.

Our problem of determining best fit parameters can be cast in terms of data-fitting problem. If we consider the mean square differences between two distribution functions, the L^2-norm of the residual can be defined as follows:

$$R = \left\|\ln p_2(\mathbf{x}|\alpha,\beta)-\ln p_1(\mathbf{x})\right\|_{L^2}^2 = \left\|\ln\frac{p_2(\mathbf{x}|\alpha,\beta)}{p_1(\mathbf{x})}\right\|_{L^2}^2 \qquad (10)$$

We do not use least square fitting function to obtain the parameters which minimize R since there is a constraint that $p_1(\mathbf{x}) > 0$ and $p_2(\mathbf{x}) > 0$, and $p_1(\mathbf{x})$ can be zero when there is no sample points. In addition, `lsqcurvefit()` function could cause an inaccurate solution if $p_1(\mathbf{x})$ is unreasonably small. Instead, we observe the relationship between the risk value R and the parameters α and β.

As mentioned in our experiments, we use ten dimensional vector. It is not possible to show graphically the sample distribution for the 10-dimensional model. Hence, instead of analyzing the multivariate model which uses ten elements, a bivariate model is analyzed as a specific and simple example of multivariate model. Each data vector of the bivariate model is a pair of a coefficient and its parent.

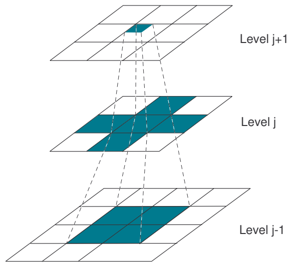

Fig. 3. Selected elements of vector $\mathbf{x}$ from wavelet coefficients in our experiments ($d = 10$).

In order to obtain proper values of the parameters (α, β) for the bivariate model, we plot the distribution of R as function of (α, β) in Fig. 2. A contouring region on the $\alpha\beta$-plane which has the minimal values of R can be chosen for determining the parameters. So the best fit parameters for the bivariate case are roughly in the contouring region bounded by $\alpha \in (0, 0.8]$ and $\beta \in [0.3, 0.8]$. The contour graphs in Fig. 2 can be different depending on the image types or databases. It is difficult to find one set of the best fit parameters for all cases. However, in our experiments we found that the parameters are not different much for different images.

For the higher-dimensional case, the best fit parameters for MGGD model vary depending on the number and type of elements (parent, neighbours, offsprings, etc.). However, we found that the best fit parameters for the bivariate case are not far from those for the 10 dimensional case which is used in our experiments with certain types of natural images.

4. EXPERIMENTAL RESULTS

In our experiments, we choose 10 elements as depicted in Fig. 3. As described in the previous section, proper parameters for the statistical model are necessary for good estimation. This could be difficult since they are decided case by case empirically depending on the type of images, the subband in wavelet domain and the chosen elements of vector $\mathbf{x}$. However, our experiments show that when the parameters are chosen inside certain range as described before, the denoising quality is very similar. For the chosen elements in our experiments, we select the parameters of MGGD model as $\alpha = 1/6$ and $\beta = 1/2$. These parameters are chosen based on Fig. 2. We use them throughout our experiments. Since the noise is independently distributed, the estimation of covariance matrix $\Sigma_{\mathbf{x}}$ for a model is given by $\widehat{\Sigma}_{\mathbf{x}} = \Sigma_{\mathbf{y}} - \sigma^2 I$, where $\Sigma_{\mathbf{y}}$ is a covariance matrix from given noisy wavelet coefficients. We use a local covariance from 7×7 neighbouring window which surrounds each element of $\mathbf{y}$. Recent works have empirically shown that the local variance is usful for image denoising [3, 6].

Tables 1 and 2 list the PSNR's of the proposed and other existing methods for two popular images, $Lena$ and $Boat$. The results are categorized in terms of the type of the wavelet used since denoising results are dependent on the wavelet transforms. Our algorithm outperforms all other methods reported in the literature when the same wavelet transform is used, in particular Daubechies 8 filter, which is the most popular scalar wavelet used for image denoising.

Fig. 4. Cropped $Lena$ image and its denoised images by some denoising methods using Daubechies 8 filter: Clean image (top-left); Noisy image (σ=20, 22.14dB, top-center); VisuShrink (26.59dB, top-right); BayesShrink (30.14dB, bottom-left); GSM (31.03dB, bottom-center); Proposed method (d=10, α=1/6, β=1/2, 31.44dB, bottom-right).

We also illustrate 128×128 size of cropped *Lena* image for visually compared evaluation in Fig. 4. It takes about 30 seconds for a 512×512 image with Daubechies 8 filter on 2.4 GHz Pentium IV PC when 10 elements are used. However, when $\hat{\mathbf{x}}$ can be calculated explicitly without solving Eq. (8), it only takes less than 3 seconds under the same condition.

In Tables 1 and 2, we have made use of the functions in `WaveLab` for *VisuShrink* and *SureShrink*. For the other existing methods, the experimental results are taken from the original papers and the PSNR table in [6]. GSM results for Daubechies 8 filters are obtained from the software offered by Portilla [4]. If an original paper shows the results in terms of mean squared error (MSE), it is converted into PSNR value by the following formula : PSNR(dB) $= 10\log_{10}\frac{255^2}{MSE}$. Also PSNR values from [6] are converted by subtracting 0.03 dB since they define PSNR(dB) $= 10\log_{10}\frac{256^2}{MSE}$.

In Section 3, we have analyzed the model and the estimated parameters α and β. The regions around the optimal values for α and β are almost flat and have gentle slopes as seen in Fig. 2. Fig. 5 shows PSNR value distribution as function of the parameters α and β by using 512×512 *Lena* image with σ=20 and d=10. The desirable parameter coordinates (α, β) lie in an elliptical narrow shape and the graph has some similarity with the residue contour graph in Fig. 2. One thing we need to consider about the parameters is that their range could slightly vary depending on the image type or image database. For example, the proper parameters from nematode images could be different from the ones from nebula images. Therefore, it might be possible to get better image denoising performance by varying the parameters depending on the cases.

5. REFERENCES

[1] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sept. 2000.

[2] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive

Approach	Wavelet	PSNR(dB)				
		σ=10	σ=15	σ=20	σ=25	σ=30
Noisy image	–	28.12	24.62	22.14	20.16	18.60
Proposed	Daub. 8	34.55	32.71	31.44	30.46	29.64
VisuShr. (soft) [11]		28.70	27.40	26.59	26.04	25.60
VisuShr. (hard) [11]		30.65	28.89	27.76	27.02	26.33
SureShr. [12]		33.42	31.50	30.17	29.18	28.47
Bi-shrink. [5]		33.91	32.03	30.70	29.78	28.91
Local Bi-Shr. [6]		34.33	32.48	31.16	30.12	29.38
HMT [13]		33.81	31.73	30.36	29.21	28.32
LAWMAP (7x7) [3]		34.24	32.27	30.92	29.90	
GSM [4]		34.23	32.35	31.03	30.23	29.21
BayesShr. [1]		33.29	31.38	30.14	29.19	28.45
AdaptShr [2]	Symm. 8		32.36	31.04	30.04	
Proposed	CWT	35.35	33.70	32.46	31.48	30.68
Local Bi-Shr. [6]		35.31	33.64	32.37	31.37	30.51
CHMT [14]		34.90			29.90	
MMSE [15]	OEB 10/18	34.93	33.01	31.69	30.60	
SI-AdaptShr [2]	SI-Symm.8		33.37	32.09	31.07	
GSM [4]	Steerable pyramid	35.61	33.90	32.66	31.69	

Table 1. Comparison table for our and other existing methods with different levels of Gaussian noise (*Lena* 512 × 512). CWT and OEB stands for dual-tree *complex wavelet transform* and *overcomplete expansion biorthogonal* filter respectively.

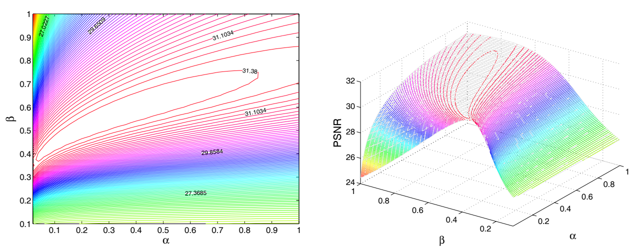

Fig. 5. PSNR contour by varying parameters α and β with d=10 for noisy $Lena$ image ($\sigma = 20$).

Approach	Wavelet	PSNR(dB)				
		σ=10	σ=15	σ=20	σ=25	σ=30
Noisy image	–	28.13	24.63	22.10	20.18	18.59
Proposed	Daub. 8	32.54	30.69	29.38	28.34	27.56
VisuShr. (soft) [11]		26.64	25.34	24.59	24.06	23.68
VisuShr. (hard) [11]		28.61	26.90	25.82	25.03	24.46
SureShr. [12]		31.83	29.88	28.55	27.50	26.73
HMT [13]		32.25	30.28	28.81	27.65	26.80
LAWMAP (7x7) [3]		32.22	30.37	28.97	27.88	27.03
Bivariate Shr. [5]		32.22	30.22	28.90	27.88	27.08
Local Bi-Shr. [6]		32.39	30.52	29.15	28.11	27.26
GSM [4]		32.39	30.41	29.03	27.99	27.15
BayesShr. [2]		31.77	29.84	28.45	27.37	26.57
Proposed	CWT	33.31	31.46	30.14	29.12	28.24
Local Bi-Shr. [6]		33.07	31.33	30.05	29.03	28.28
GSM [4]	Steerable pyramid	33.58	31.70	30.38	29.37	

Table 2. Comparison table for our and other existing methods with different levels of Gaussian noise ($Boat$ 512 × 512).

wavelet thresholding with context modeling for image denoising," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1522–1531, Sept. 2000.

[3] M. K. Mihcak, K. Ramchandran I. Kozintsev, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, 1999.

[4] J. Portilla, M. Wainwright V. Strela, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.

[5] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, 2002.

[6] L. Sendur and I. W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, 2002.

[7] G. Y. Chen and T. D. Bui, "Multiwavelet denoising using neighbouring coefficients," *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 211–214, 2003.

[8] T. T. Cai and B. W. Silverman, "Incorporating information on neighbouring coefficients into wavelet estimation," *Sankhya: The Indian Journal of Statistics, Serie A*, vol. 63, pp. 127–148, 2001.

[9] S. G. Mallat, "A theory for multiresolution signal decomposition:the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[10] C. Fernández, J.Osiewalski, and M. Steel, "Modelling and inference with v-spherical distributions," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1331–1340, 1995.

[11] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, pp. 613–627, 1995.

[12] D. L. Donoho and I. M. Johnstone, "Adaptating to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[13] M. S. Crouse, R. D Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden markov models," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, 1998.

[14] H. Choi, J. K. Romberg, R. G. Baraniuk, and N. G. Kingsbury, "Hidden markov tree modeling of complex wavelet transforms," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, June 2000, vol. 1, pp. 133–136.

[15] X. Li and M. T. Orchard, "Spatially adaptive image denosing under overcomplete expansion," in *IEEE International Conference on Image Processing (ICIP)*, Sept. 2000, vol. 3, pp. 300–303.