Journal of Zhejiang University SCIENCE A
ISSN 1673-565X (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

# Mean shift based log-Gabor wavelet image coding

LI Ji-liang†, FANG Xiang-zhong, HOU Jun

(*Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China*)

†E-mail: jilianglee@163.com

Received Sept. 13, 2006; revision accepted Dec. 18, 2006

**Abstract:** In this paper, we propose a sparse overcomplete image approximation method based on the ideas of overcomplete log-Gabor wavelet, mean shift and energy concentration. The proposed approximation method selects the necessary wavelet coefficients with a mean shift based algorithm, and concentrates energy on the selected coefficients. It can sparsely approximate the original image, and converges faster than the existing local competition based method. Then, we propose a new compression scheme based on the above approximation method. The scheme has compression performance similar to JPEG 2000. The images decoded with the proposed compression scheme appear more pleasant to the human eyes than those with JPEG 2000.

**Key words:** Sparse approximation, Log-Gabor, Image coding, Mean shift, Overcomplete
**doi:**10.1631/jzus.2007.A0620 **Document code:** A **CLC number:** TN919.8

## INTRODUCTION

State-of-the-art image compression is based on orthogonal transform, such as discrete cosine transform and a series of dyadic wavelet transforms. However, those schemes suffer some severe limitations. Wavelets for example fail to capture the regularities of contours since they cannot sparsely represent the 1D singularities of 2D signals (Peotta *et al*., 2006). Biorthogonal wavelets are not optimal and artifacts can be introduced by the compression (Fischer *et al*., 2006). A lot of research effort was aimed at coding a natural image more efficiently from different approaches (Jiang *et al*., 2005; Le Pennec and Mallat, 2005; Peotta *et al*., 2006; Wakin *et al*., 2006). One of those approaches is based on sparse representation with an overcomplete dictionary (Mallat and Zhang, 1993). This approach can be more efficient than those based on orthogonal transforms in terms of entropy (Kreutz-Delgado, 2003), and is the strategy used by the visual cortex for representing images (Olshausen and Field, 1997). One of the key steps of sparse representation is selecting subdictionary. Many subdictionary selection methods (Mallat and Zhang, 1993; Kreutz-Delgado, 2003; Zhang *et al*., 2005) were proposed in the past, and much progress was made in these works. Nevertheless, minimizing the size of the subdictionary is still an unsolved problem (Pece, 2002).

The biologically plausible sparse representation method proposed in (Fischer *et al*., 2006) uses overcomplete log-Gabor wavelets for decomposition, where the log-Gabor function is a model for cortex cell (Field, 1987). A local competition based algorithm is employed to select the local maximum coefficients, and to concentrate the energy on the selected coefficients. A coefficient is the projection of the image on a wavelet, and the set of overcomplete log-Gabor wavelets can be regarded as an overcomplete dictionary. So coefficient selection is equivalent to subdictionary selection.

In this paper, we propose a sparse overcomplete approximation method based on the ideas of overcomplete log-Gabor wavelet decomposition, mean shift (Comaniciu and Meer, 2002) and energy concentration. A mean shift based algorithm is proposed to select the necessary coefficients for rapid searching. Then all the energy of the coefficients is concentrated on the selected coefficients with an algorithm similar to that in (Fischer *et al*., 2006). This approximation method can sparsely approximate images, and has less computational cost than the method in (Fischer *et al.*, 2006). Following this approximation method, a compression scheme is proposed. This scheme employs the proposed sparse overcomplete approximation method, and has compression performance similar to JPEG 2000. Images decoded with the proposed compression scheme appear visually more pleasant to the human eyes than those with JPEG 2000.

This paper is organized as follows. Section 2 introduces the new sparse overcomplete log-Gabor wavelet approximation method. Section 3 shows our compression scheme. Experiments and analysis are given in Section 4. Section 5 concludes the paper.

## A SPARSE OVERCOMPLETE APPROXIMATION METHOD

Our sparse overcomplete approximation method consists of overcomplete log-Gabor wavelets decomposition, mean shift based coefficient selection, and energy concentration. Details are described in what follows.

### Overcomplete log-Gabor wavelet transform

We use the same filter function as that in (Fischer *et al.*, 2006) to perform decomposition, because of the log-Gabor function's close similarity to the cortex cell model. To get good results, four scales and eight orientations are used. All filters have one octave-scale bandwidth and π/8 angular bandwidth. An additional low pass filter is applied to transmit the luminance, and a high pass filter allows encoding the highest frequencies. Each filter's output is called a channel. The real parts of the channels contain enough information to reconstruct the original image (http://www.csse.uwa.edu.au/~pk/research/matlabfns/PhaseCongruency/Docs/convexpl.html). We use 34 real channels while the method in (Fischer *et al.*, 2006) uses 16 complex channels and 2 real channels. All channels are downsampled without alias. The downsampled coefficients pyramid is denoted by $\boldsymbol{h}_1$. Assume $\boldsymbol{I}$ is the original image, $T$ and $T^{\dagger}$ are the wavelet transform and inverse transform respectively. Then $\boldsymbol{h}_1=T(\boldsymbol{I})$, and $\boldsymbol{I}=T^{\dagger}(\boldsymbol{h}_1)$.

### Mean shift based coefficient selection algorithm

Because mean shift is a simple procedure for searching for local maximums (Cheng, 1995), we apply it to search the important coefficients, i.e., the coefficients with local maximal values. The coefficient selection algorithm is based on the mean shift. The details of the mean shift procedure are referred to (Cheng, 1995; Comaniciu and Meer, 2002). Only necessary definitions are given here.

Let $X$ be the $d$-dimensional Euclidean space $\mathbb{R}^d$, and $x^i$ denote the $i$th component of $\boldsymbol{x}$, where $\boldsymbol{x}\in X$. The norm of $\boldsymbol{x}$ is a nonnegative number $\|\boldsymbol{x}\|$, such that

$$\|\boldsymbol{x}\|^2=\sum_{i=1}^{d}|x^i|^2. \tag{1}$$

Then the kernel used in this study is

$$G_{B_\mathrm{s},B_\mathrm{r}}(\boldsymbol{x})=\frac{C}{B_\mathrm{s}^2 B_\mathrm{r}^2}g\left(\|\boldsymbol{x}^\mathrm{s}/B_\mathrm{s}\|^2\right)g\left(\|\boldsymbol{x}^\mathrm{r}/B_\mathrm{r}\|^2\right),$$

where $g(\cdot)$ is the Gaussian kernel's profile (Cheng, 1995), $\boldsymbol{x}^\mathrm{s}$ is the position of $\boldsymbol{x}$ in the corresponding subband, $\boldsymbol{x}^\mathrm{r}$ is the value of $\boldsymbol{x}$, $B_\mathrm{s}$ and $B_\mathrm{r}$ are the bandwidth parameters associated with $\boldsymbol{x}^\mathrm{s}$ and $\boldsymbol{x}^\mathrm{r}$ respectively, and $C$ is the corresponding normalization constant.

The values of $B_\mathrm{s}$ and $B_\mathrm{r}$ are constants related to the character of images. The experiential ranges are $B_\mathrm{s}\in[3, 10]$ and $B_\mathrm{r}\in[2, 5]$.

Let $p$ and $q$ denote two points on the straight line $\overleftrightarrow{pq}$, and $\theta$ denote filter direction. The line $\overleftrightarrow{pq}$ is approximately vertical to the filter direction if

$$\frac{p|_y-q|_y}{p|_x-q|_x}\in\begin{cases}\left(-\infty,\cot(-\pi/16)\right]\cup\left(\cot(\pi/16),+\infty\right), & \theta=0\\ \left[\cot(\theta+\pi/16),\cot(\theta-\pi/16)\right), & \text{otherwise}\end{cases}$$

where $p|_y$ is the $y$ coordinate of $p$, and $p|_x$ is the $x$ coordinate. As $\theta$ is known, the cotangent functions $\cot[\theta\pm(\pi/16)]$ can be calculated in advance.

All transform coefficients in $\boldsymbol{h}_1$ are zeroed out if they are below a fixed threshold $\Gamma$. This reduced version of $\boldsymbol{h}_1$ is denoted by the set $\{\Theta_J\}_{J=1,2,\ldots}$, where $\Theta_1$ and $\Theta_2$ denote the low frequency channel and the high frequency channel respectively, and $\{\Theta_J\}_{J=3,4,\ldots}$ denote other channels.

The coefficient selection method is described as follows:

Step 1: Run the mean shift procedure on each $\Theta_J$, where $J$=1, 2, …, and use $\{z_i^J\}_{i=1,2,\ldots}$ to denote the positions of convergence points in the corresponding channel.

Step 2: (1) For channels $\{\Theta_J\}_{J=3,4,\ldots}$:

For any $z_i^J$ and $z_m^J$, if (i) they are closer than $B_s^l$ in distance, (ii) they are closer than $B_r^l$ in value, (iii) the straight line $\overleftrightarrow{z_j^J z_m^J}$ is approximately vertical to the filter direction, concatenate $z_j^J$ and $z_m^J$. The positions of the coefficients located on the concatenating lines are denoted by $\{\overline{z}_i^J\}_{i=1,2,\ldots}$.

(2) For channels $\{\Theta_J\}_{J=1,2}$:

$$\{\overline{z}_i^J\}_{i=1,2,\ldots} = \{z_i^J\}_{i=1,2,\ldots}.$$

The coefficients located on $\{\overline{z}_i^J\}_{i=1,2,\ldots}$ are the selected coefficients.

### Iterative energy concentration

Assume $\{\boldsymbol{h}_j\}_{j=2,3,\ldots}$, $\{\boldsymbol{a}_j\}_{j=2,3,\ldots}$, and $\{\boldsymbol{r}_j\}_{j=2,3,\ldots}$ are three coefficients pyramids series. The energy concentration is implemented by performing the following procedures repeatedly until the iteration times reach the given times:

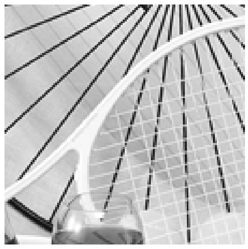
$$\boldsymbol{a}_j(k) = \begin{cases} \boldsymbol{h}_{j-1}(k), & k \in \{\overline{z}_i^J\}_{i=1,2,\ldots}, \\ 0, & \text{otherwise}, \end{cases}$$

$$\boldsymbol{r}_j = \boldsymbol{h}_{j-1} - \boldsymbol{a}_j,$$

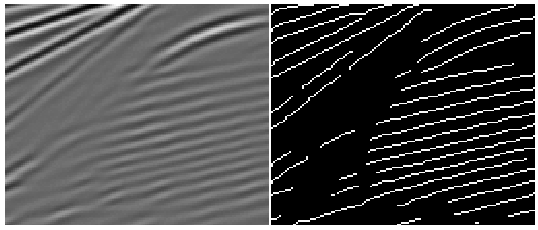
$$\boldsymbol{h}_j = \boldsymbol{a}_j + T(T^{\dagger}(\boldsymbol{r}_j)).$$

After the above concentration, most of the energy is concentrated on the selected coefficients. The selected coefficients are used to approximate the original image.

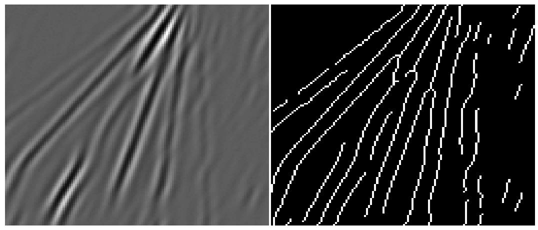
Some results of the above overcomplete wavelet approximation method are given in Fig.1. Fig.1a is a part of the original image. Fig.1b and Fig.1c are two illustrations of the above method, where the left parts of both images are transform coefficients, and the right parts are the corresponding concentration results. From Fig.1, we can see that the proposed method significantly reduces the quantity of coefficients.



(a)



(b)



(c)

**Fig.1 Illustrations of the proposed approximation method. (a) Part of the original image; (b) The second-scale, (5π/8)-orientation channel; (c) The second-scale, (7π/8)-orientation channel**

## COMPRESSION SCHEME

We apply the above proposed sparse overcomplete approximation method to a new compression scheme. The total compression scheme is shown in Fig.2. We employ the quantization methods in (Zeng *et al*., 2000) to quantize the coefficients, with the related quantization parameters taking the recommended values in (Zeng *et al*., 2000). JPEG 2000 is the entropy coder (ISO/IEC N1878, 2000).

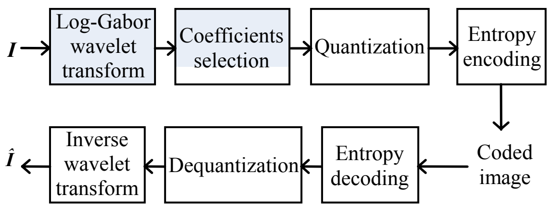



**Fig.2 The compression scheme. *I* denotes input image; $\hat{I}$ means reconstructed image**

## EXPERIMENTS AND RESULTS

We test the performance of the proposed compression scheme from two approaches: comparing the subjective quality of the reconstructed images and comparing the compression ratios.

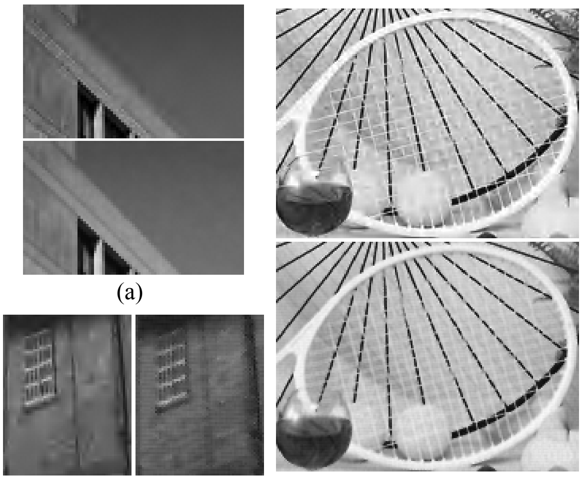
We code the same images with the proposed scheme and JPEG 2000 lossy coding respectively. The reconstructed images with similar Peak Signal-to-Noise Ratio (PSNR) are used to compare the subjective quality. For an 8-bit depth gray level image, the PSNR is defined as

$$PSNR = 10\lg\left(255\times 255/\|\boldsymbol{e}\|^2\right),$$

where $\boldsymbol{e}$ denotes the error vector and $\|\cdot\|^2$ is defined by Eq.(1). Fig.3 gives three illustrations of subjective quality comparison, which are based on three gray level images: Bank, Bike and Goldhill. Figs.3a, 3b and 3c show the performance comparison of the Bank, Goldhill and Bike, respectively. The images decoded with the proposed scheme look smoother and have less high frequency artifacts, such as blocking artifacts, than those with JPEG 2000. The edges quality is similar to that in JPEG 2000 decoded images. On the whole, the images decoded with the proposed scheme appear more pleasant to human eyes. The smoothness and fewer artifacts are caused by the shape of log-Gabor filters, and the log-Gabor function is a model for cortex cell (Field, 1987). The edge's quality is preserved by the proposed coefficient selection algorithm and the energy concentration algorithm. Those two algorithms enhance locally maximal coefficients as well as restrain other coefficients, which is similar to the masking effect (Legge and Foley, 1980) of human visual system. The above close similarity to biological models helps to make the decoded images pleasant to human eyes.

Table 1 gives some of the comparison of compression performance. All the experiments based on 23 images show that the images encoded with the proposed scheme are about 4% larger than those with JPEG 2000, i.e., the compression performance is similar to that of the JPEG 2000 lossy coding.

The computation cost of one energy concentration iteration is $O(N\log N)$ (Fischer *et al.*, 2006), where $N$ is the number of input pixels. The complexity of the mean shift based coefficient selection algorithm is $O(k\log k)$ (Comaniciu and Meer, 2002), where $k$ is the number of the nonzero transform coefficients in a channel. The computational cost of the mean shift based coefficient selection algorithm is far less than that of the energy concentration algorithm because $k \ll N$. So the complexity of the sparse overcomplete log-Gabor wavelet approximation method is $O(N\log N)$.



(a) (b) (c)

**Fig.3 Comparison between the proposed method and JPEG 2000. (a) Parts of reconstructed Bank image. Top: coded with JPEG 2000 (*PSNR*=23.60). Bottom: coded with the proposed compression scheme (*PSNR*=23.51); (b) Parts of reconstructed Goldhill image. Left: coded with JPEG 2000 (*PSNR*=30.86). Right: coded with the proposed compression scheme (*PSNR*=31.02); (c) Parts of reconstructed Bike image. Top: coded with JPEG 2000 (*PSNR*=31.25). Bottom: coded with the proposed compression scheme (*PSNR*=31.08)**

**Table 1 The comparison of compression performance**

| Image | JPEG 2000 | | The proposed scheme | |
|---|---|---|---|---|
| | *PSNR* (dB) | Bit rate (bpp) | *PSNR* (dB) | Bit rate (bpp) |
| Bike | 40.00 | 1.94 | 40.05 | 1.99 |
| Goldhill | 40.01 | 1.65 | 39.98 | 1.86 |
| Bank | 40.05 | 1.58 | 39.90 | 1.57 |
| Cameraman | 39.98 | 1.73 | 40.09 | 1.81 |
| House | 40.11 | 0.89 | 39.93 | 0.90 |

bpp: bits per pixel

The proposed approximation method saves about 18%~25% iterations compared with the local competition based method in (Fischer *et al.*, 2006).

CONCLUSION

In this study, we propose a new sparse approximation method and propose a new compression scheme. The approximation method is based on the ideas of overcomplete log-Gabor wavelet decomposition, mean shift, and energy concentration. It significantly reduces the redundancy in log-Gabor wavelet coefficients, and converges faster than the local competition method in (Fischer *et al*., 2006). The proposed compression scheme has compression ratio similar to JPEG 2000, which is one of the most efficient standards. The particular advantage of the proposed scheme is its close similarity to biological models. This advantage makes the decoded images look smoother, have fewer artifacts, and appear more pleasant to the human eyes than those decoded with JPEG 2000.

Our further study aims at finding an adaptive method to determine the related parameters, such as the threshold $\Gamma$, the bandwidth parameters $B_s$ and $B_r$.

**References**

Cheng, Y., 1995. Mean shift, mode seeking, and clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **17**(8):790-799. [doi:10.1109/34.400568]

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **24**(5):603-619. [doi:10.1109/34.1000236]

Field, D.J., 1987. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Amer. A*, **4**(12):2379-2394.

Fischer, S., Cristóbal, G., Redondo, R., 2006. Sparse overcomplete Gabor wavelet representation based on local competitions. *IEEE Trans. on Image Processing*, **15**(2):265-272. [doi:10.1109/TIP.2005.860614]

ISO/IEC N1878, 2000. JPEG2000 Verification Model 8.5.

Jiang, Y.X., Zhong, Z.F., Wang, L.W., 2005. A Wavelet Coding Preprocessing Algorithm Based on Bayesian Estimation for Image Compression. Proc. 4th International Conference on Machine Learning and Cybernetics. Guangzhou, p.5467-5472. [doi:10.1109/ICMLC.2005.1527910]

Kreutz-Delgado, K., 2003. Dictionary learning algorithms for sparse representation. *Neural Computation*, **15**(2):349-396. [doi:10.1162/089976603762552951]

Legge, G., Foley, J., 1980. Contrast masking in human vision. *J. Math. Imaging and Vision*, **12**:1458-1471.

Le Pennec, E., Mallat, S., 2005. Sparse geometric image representations with bandelets. *IEEE Trans. on Image Processing*, **14**(4):423-438. [doi:10.1109/TIP.2005.843753]

Mallat, S.G., Zhang, Z., 1993. Matching pursuit in a time-frequency dictionary. *IEEE Trans. on Signal Processing*, **41**(12):3397-3415. [doi:10.1109/78.258082]

Olshausen, B.A., Field, D.J., 1997. Sparse coding with an overcomplete basis set: a strategy employed by Vl? *Vision Research*, **37**:3311-3325. [doi:10.1016/S0042-6989(97)00169-7]

Pece, A.E.C., 2002. The problem of sparse image coding. *J. Math. Imaging and Vision*, **17**:89-108. [doi:10.1023/A:1020677318841]

Peotta, L., Granai, L., Vandergheynst, P., 2006. Image compression using an edge adapted redundant dictionary and wavelets. *Signal Processing*, **86**:444-456. [doi:10.1016/j.sigpro.2005.05.023]

Wakin, M.B., Romberg, J.K., Choi, H., Baraniuk, R.G., 2006. Wavelet-domain approximation and compression of piecewise smooth images. *IEEE Trans. on Image Processing*, **15**(5):1071-1087. [doi:10.1109/TIP.2005.864175]

Zeng, W., Daly, S., Lei, S., 2000. Point-Wise Extended Visual Masking for JPEG-2000 Image Compression. Proc. IEEE International Conference on Image Processing. Vancouver, CA, **1**:657-660.

Zhang, L., Bao, P., Wu, X.L., 2005. Multiscale LMMSE-based image denoising with optimal wavelet selection. *IEEE Trans. on Circuits and Systems for Video Technology*, **15**(4):469-481. [doi:10.1109/TCSVT.2005.844456]