WARPLETS: AN IMAGE-DEPENDENT WAVELET REPRESENTATION

Abhir Bhalerao and Roland Wilson

Department of Computer Science, University of Warwick, Coventry, UK
{abhir|rgw}@dcs.warwick.ac.uk

ABSTRACT

A novel image-dependent representation, warplets, based on self-similarity of regions is introduced. The representation is well suited to the description and segmentation of images containing textures and oriented patterns, such as fingerprints. An affine model of an image as a collection of self-similar image blocks is developed and it is shown how textured regions can be represented by a single prototype block together with a set of transformation coefficients. Images regions are alligned to a set of dictionary blocks and their variability captured by PCA analysis. The block-to-block transformations are found by Gaussian mixture modelling of the block spectra and a least-squares estimation. Clustering in the Warplet domain can be used to determine a warplet dictionary. Experimental results on a variety of images demonstrate the potential of the use of warplets for segmentation and coding.

1. INTRODUCTION

Many natural scenes, biological images, or images containing repetitions, often have a tiling of some image 'patch' or texton. Image models that have exploited image self-similarity for image analysis are the fractal image coding methods of Jacquin and others [1], and warping methods first reported by Volet [2] for texture. Syntactical or structural approaches attempt to capture the repetition and self-similarity of a texture by picking a texton and defining rules for its placement to synthesize the entire texture. For some types of texture, a structural model alone is unsuitable, so statistical texture models, such as MRFs, have been used to parameterize the randomness of the texture. With combined structural and random models [3] a stochastic element is 'added' to a purely deterministic local texton model to account for the effects of texture variation due to surface properties, lighting or noise in the imaging process. Natural textures, however, differ fundamentally from man-made textures by a degree of randomness, syntactical errors or faults, that is inherent at the structural level. A way to compactly represent such imagery for the purposes of segmentation and content-based image retrieval is desirable.

Using Gaussian Mixture Modelling (GMM), we build on the work of Hsu, Calway and Wilson [4, 5], whose two-component affine texture model combined a structural texton plus stochastic residual that enabled the synthesis of a range of textures from periodic to random. Moreover, this affine placement model itself can accurately model small structural variations which seem important in natural texture. All image blocks are affine-warped to the coordinate frame of a representative texton and a PCA is done. The mean plus modes of the eigen image analysis, and the set of block-to-block transformations make up our warplet representation. Feature clustering is used to determine a representative set of prototypes for the image. We thus obtain a concise way to describe the image and perform image segmentation using image features which are small representative patches from the image. Clustering is invariant to local affine transformations of these patches, and the accuracy of the representation can be measured by image reconstruction.

2. IMAGE WARPLETS

Image regions are modeled as sets of repeating texture patches, assuming that each region can be adequately represented by affine coordinate transformations of some prototypical patch (the source texton) and an intensity scaling. Denoting the coordinates of a target patch, f_j, by $\mathbf{y} = (p, q)^T, 0 \leq p, q, \leq N$, the target patch estimate is

$$\hat{f}_j(\mathbf{y}) = \alpha_{ij} f_i(T_{ij}(\mathbf{x})), \qquad T_{ij}(\mathbf{x}) = A_{ij}\mathbf{x} + \mathbf{t}_{ij} \quad (1)$$

$$\alpha_{ij}^2 = \sum_{\mathbf{x}} f_i^2(\mathbf{x}) / \sum_{\mathbf{x}} f_j^2(\mathbf{x}) \quad (2)$$

where T_{ij} is a 2D affine transformation of the coordinates of some source patch f_i and α_{ij} is a scaling for image intensity changes. The transformation consists of a linear part, A_{ij}, and a translation, $\mathbf{t}_{ij}$. It is convenient to use overlapping square image patches (blocks) as warplets of size B, such as 16×16, 32×32 etc. By using a squared cosine window function centred on the block, $w(\mathbf{y}) = \cos^2[\pi p/B]\cos^2[\pi q/B]$ and having the blocks overlap by 50%, it is possible to sum the transformed patches, $w(\mathbf{y})f_j(\mathbf{y})$, without blocking artifacts being introduced.

To estimate the affine transformations T_{ij}, the Fourier transform of the source patch, $F_i(\mathbf{u})$, is used to separate the transformation,

$$F_j(\mathbf{u}) = \frac{1}{|detA_{ij}|}\exp(i\mathbf{u}^T\mathbf{t}_{ij})|F_i(|A_{ij}^T|^{-1}\mathbf{u})| \quad (3)$$

such that the linear part, A_{ij}, affects only the amplitude spectrum and the the translation $\mathbf{t}_{ij}$ is exhibited as a phase gradient. The amplitude spectrum, $|F_i(\mathbf{u})|$, is modelled as a two-dimensional, M component Gaussian mixture

$$G_i(\mathbf{u}) = \sum_m^M a_m \exp(-\mathbf{u}^T C_m^{-1}\mathbf{u}/2), \quad (4)$$

having centroids fixed on the origin of the spectrum and coordinates $\mathbf{u} = (r, s)^T$. The Gaussian parameters, $\{a_m, C_m\}$ are then estimated by minimizing the residual error,

$$\sum_{\mathbf{u}}(|F_i(\mathbf{u})| - G_i(\mathbf{u}; a_m, C_m))^2, \quad (5)$$

using a standard non-linear optimization method, Levenberg-Marquardt (LM). LM is a conjugate gradient descent method which requires the gradient with respect to the fitting parameters to be known:

$$\frac{dG_i(\mathbf{u})}{da_m} = G_i(\mathbf{u})/a_m, \qquad \frac{dG_i(\mathbf{u})}{dC_m^{-1}} = G_i(\mathbf{u})\mathbf{u}\mathbf{u}^T/2. \quad (6)$$

In the experiments presented below, we have found that a mixture with 2 components is sufficient to model the general shape and amplitude of oriented and periodic patches so that a linear transformation of the model can be *uniquely* fit to the amplitude spectrum of a target image block.

The second step of the affine transformation estimation uses the mixture model of the source block spectrum and searches for a linear transformation that minimizes the squared residuals, $\sum_{\mathbf{u}}(G_i(A_{ij}(\mathbf{u})) - |F_j(\mathbf{u})|)^2$. This search can be again performed by the LM method. This time the gradients of $H_i(\mathbf{u}) = G_i(A(\mathbf{u}))$ w.r.t. the parameters of the linear transformation matrix are needed:

$$\frac{dH_i(\mathbf{u})}{dA} = -H_i(\mathbf{u})\sum_m^M C_m^{-1}A\mathbf{u}\mathbf{u}^T \quad (7)$$

The final step is to estimating the translation, t_{ij}, which is exhibited in the phase spectrum of the block DFT. An estimate of a transformed block $\hat{f}_j(\mathbf{y})$, can be synthesized by the applying the linear transformation: $\mathbf{y} = A_{ij}\mathbf{x}$. The translation, $\mathbf{t}_{ij}$, is then taken as the peak location of the cross-correlation i.e. $\arg\max_{\mathbf{t}}[\hat{f}_j \star f_j]$.

The Image Warplet, $\mathcal{W}_i$, is defined as the set of all image blocks j transformed to the coordinate frame of block i using T_{ij}^{-1},

$$\mathcal{W}_i = \{w_{ij}(\mathbf{x}) = \hat{f}_j^i(\mathbf{x}), \mathbf{x} = A_{ij}^{-1}(\mathbf{y} - \mathbf{t}_{ij})\forall j\} \quad (8)$$

PCA (or harmonic analysis) of the warplet blocks, w_{ij} is used to encode their variability. Then, any image block is represented by the mean warplet plus an appropriate number of modes of variation in the Warplet domain, and the corresponding block-to-block affine warps, T_{ij}.

3. SELF-SIMILARITY CLUSTERING

To discover a set of prototypical warplets, ideally one for each image region that appropriately ‘span’ the image, a clustering approach is used. In the supervised case, we cluster not just to the representative (given) cluster centroids, but to an *extended family* of cluster centres which are created by taking a finite sub-group from the group $(A_{ij}, \mathbf{t}_{ij})f_l$, where l is a given prototypical block. As with the GMM model estimation described above, we extrapolate the sub-group of rotations and scaling from the amplitude spectrum prototype block. In our experiments we have restricted these prototype transformations to a sub-group of rotations and scalings of f_l and ignored shears. We have used 8 rotations and 8 scalings to create a cluster family $f_{lk}, 0 \leq k < 64$. Having applied each transformation, k, these prototype blocks give B^2 dimensional vectors: $f_{lk}(\mathbf{x}) \rightarrow \mathbf{f}_{lk}$. The remaining image blocks are then clustered using a nearest neighbour criterion. We can imagine that the data feature vectors cluster closest points on manifolds represented by the k rotation/scaling family of each prototype. Thus each data block is labelled as the closest family member which have already taken into account the affine group invariance. Schölkopf has suggested this same idea but for extending SVM vectors [6].

Unsupervised clustering can take place directly in the Warplet domain on the warped blocks, w_{ij} since the warping will have eliminated the local shape variability and allow the data vectors w_{ij} to be directly compared. A natural extension to modelling the class variability, is to perform a set of local PCA's on each resulting cluster.

4. EXPERIMENTAL RESULTS

Figure 1 shows results of self-similarity image modelling and reconstruction on a fingerprint and two wild-life images with zebras and a jaguar, and the Lena image. The mixture modelling analysis was performed on pre-whitened (Laplacian pyramid) version of the original (see [4] for details) and coarse version of the low-pass information was added back at the end. The number of parameters per pixel for these images is related to the block size, B and the image size N and can be estimated by: $P(B) = \frac{11}{(B/2)^2} + (B/N)^2$. This assumes that each affine transformation requires 6 parameters, with 1 for the intensity scaling, α, and the low-pass information encoded at twice the spatial resolution as the number of blocks in each dimension i.e. 4 extra parameters per block: the total being $6 + 4 + 1 = 11$ per block. In addition, each prototype itself has B^2 pixels per image:

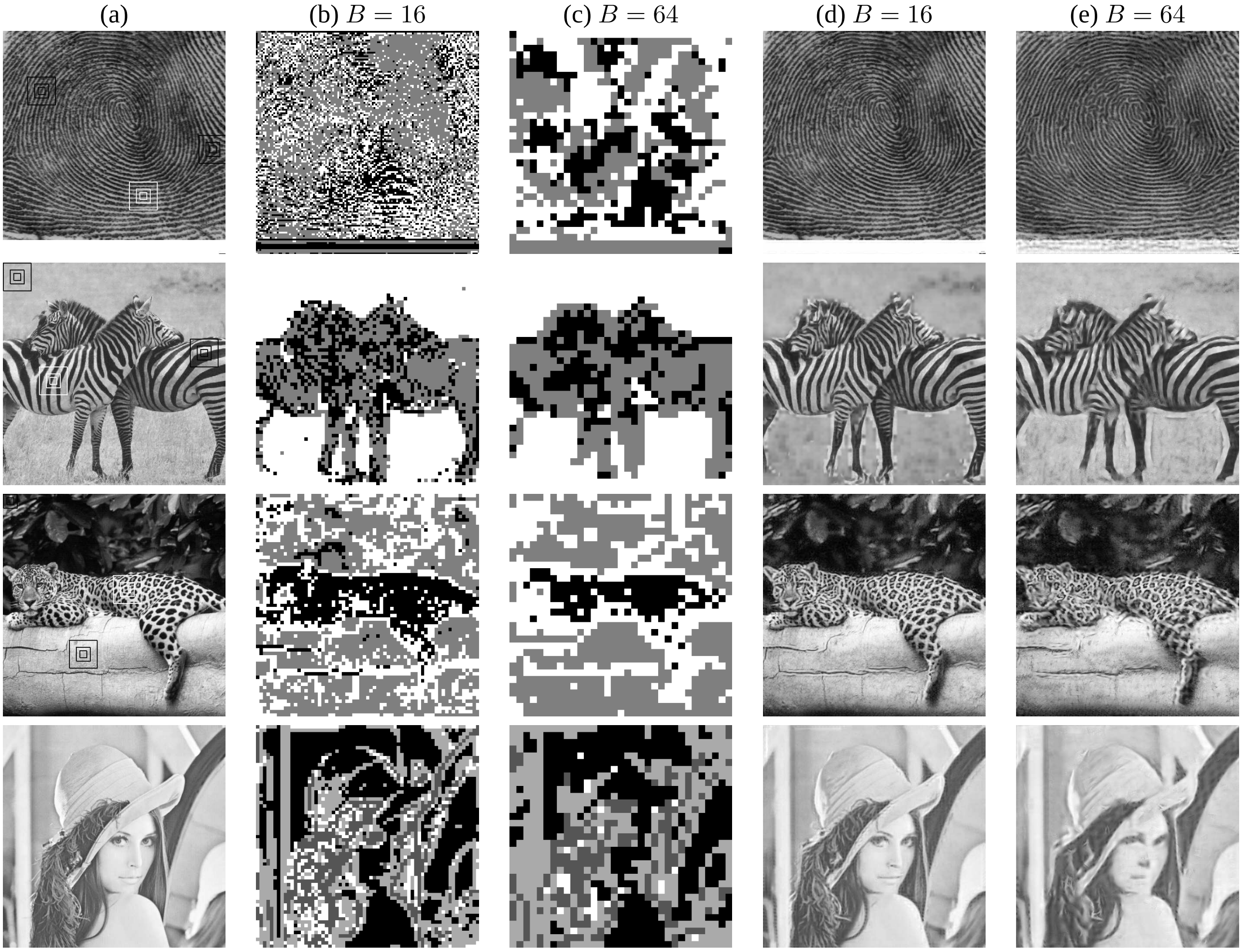


Fig. 1. **Supervised self-similarity clustering results**: (b)-(c): results on fingerprint, zebras, jaguar and Lena images grouping image blocks ($B = 16, 64$) that are related by an affine-group transformation. (d)-(e) Image reconstructions using 3 prototype texton blocks selected by self-similarity clustering results using (a).

Params/pixel	$B = 8$	$B = 16$	$B = 32$	$B = 64$
$P(B)$	0.688	0.173	0.047	0.026

The results demonstrate that the chosen warplet is able to describe and reconstruct well the original data including the cuts and minutiae of the fingerprint and the multi-directional and bifurcating stripes of the zebra. Even the reconstructions at the largest size (the lowest 'rate', at $P(64)$), exhibit the essential structural features of the image, yielding a 'painted' effect. In the zebra reconstructions, the background has no obvious deterministic component and is modelled largely by the amplitude scaling α_{ij} of (2). Where there is a structural texture present such as in the jaguar image the prototype, taken from the back of the jaguar containing the jaguar 'print', can be transformed adequately to become the leaves of the forest behind the animal as well as the bark of the log. Different blocks sizes were used to perform the clustering and subsequent reconstructions: $B = 16, 64$ results are shown in figure 1 for the fingerprint, zebra, jaguar and Lena images. In all cases, 3 warplet blocks were used to seed the cluster families (shown on the original images in figure 1, (a)). For the Lena image, blocks in the background, on the edges of the mirror and hair, and in the feathers were used. The clustering results on the fingerprint show a grouping of region blocks at orientations radially around the central whorl feature. This is to be expected, as the cluster manifold is expanded to include 8 rotations around the circle. Since warplets with horizontal, vertical and diagonal ridges were chosen, the labels roughly alternate between these three classes. In the case of the zebra, the background is clearly labelled as one group and the vertical and horizontal stripe labels together identify the two zebras. For the jaguar image, the animal is fairly well represented from the log (foreground) and the background leaves and trees. The final two columns are the corresponding reconstruction results using all three warplet both images. For the zebras, the background is well reconstructed by the texton

	PSNR dB					
Block size	16		32		64	
Params/pixel	0.688		0.173		0.047	
finger	29.0	28.7	24.6	24.8	21.2	21.5
zebra	27.6	28.0	23.6	(21.4)	19.9	20.4
jaguar	25.9	26.2	21.5	21.7	17.8	18.3
Lena	28.7	32.2	25.0	28.6	22.7	24.8

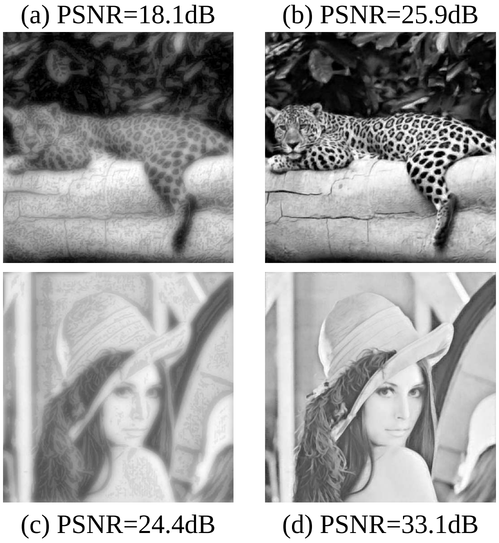
Table 1. **Estimates of reconstruction quality**. Results based on 1 or more chosen prototypes using affine invariant block based texture model. The PSNR in first column for each given block size (e.g. 16×16) is for single prototype reconstructions. The second column shows the errors (again for each block size), for multi-prototype reconstruction using the self-similarity clustering, figure 1.

representing grass but the reconstructions show some amplitude modelling errors. This is partly caused by the user specified prototype block on the back of the right-hand zebra falling on a white stripe: it does not contain any black to properly reconstruct the belly of the right-hand zebra. With the Lena image, the visual results are hard to interpret as it is difficult to manually select a representative compact set of prototypes. Table 1 presents a quantitative comparison of the single and multiple prototype reconstructions (guided by the self-similarity clustering). These results confirm the qualitative findings that the reconstructions based on the clustering are marginally better.

Image Warplet reconstruction results are presented in figure 2. The results show that the eigen-mode reconstructions of the warplet blocks are markedly better in quality with a little as 6% of the modes (15 out of a possible of $16^2 = 256$). In this way, it is possible to code the image by a set of warplets plus variations from the mean warplet block.

5. CONCLUSIONS

A new image-dependent wavelet image model has been introduced which can be used to synthesize image regions from representative image patches with good accuracy. It is particularly adept at compactly representing regions containing natural textures but can be shown to be work on other image types. The Image Warplet is built by searching for affine transformations of each image block to a given representative block or texton and then performing a PCA or harmonic analysis. The image is reconstructed by projecting all transformed blocks on to the PCA model and applying the inverse affine-transformations. We have shown how a clustering approach can be used to discover, i.e. segment, these prototypical image patches which can be used to construct a warplet dictionary for the image. The crux of supervised method is the underlying affine-invariant group model of the patch that allows us to extend the given cluster centre to its immediate affine-group family simplifying the clustering process. Unsupervised clustering can take place directly in the Warplet domain. Preliminary results are encouraging and the affine-group mixture modelling together with the wavelet self-similarity reconstruction may provide the applications such image segmentation, content based image retrieval and data compression.



(a) PSNR=18.1dB
(b) PSNR=25.9dB
(c) PSNR=24.4dB
(d) PSNR=33.1dB

Fig. 2. **Warplet reconstructions ($B = 16$)**: (a),(c) using mean warplet, (b),(d) using mean plus 15 eigen-modes. Each additional eigen-mode requires $16^2/(512^2)$ additional parameter/pixel.

6. REFERENCES

[1] A. Jacquin, "Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations," *IEEE Trans. Image Proc.*, vol. 1, pp. 18–30, 1991.

[2] P. Volet and M. Kunt, "Synthesis of Natural Structured Textures," in *Signal Processing III: Theories and Applications*, pp. 913–916. 1986.

[3] F. Liu and R. W. Picard, "Periodicity, Directionality, and Randomness: Wold Features for Image Modeling and Retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 722–733, 1996.

[4] T-I Hsu and R. Wilson, "A Two-Component Model of Texture for Analysis and Synthesis," *IEEE Trans. on Image Processsing*, vol. 7, no. 10, pp. 1466–1476, October 1998.

[5] A. D. Calway, "Image Representation Based on the Affine Symmetry Group," in *Proc. ICIP 1996*, 1996, pp. 189–192.

[6] B. Schölkopf and C. Burges and V. Vapnik, "Incorporating Invariances in Support Vector Learning Machines," in *"Artificial Neural Networks, ICANN'96"*, 1996, vol. 1112, pp. 47–52.