An Affine Symmetric Image Model and its Applications

Heechan Park, *Member, IEEE,* Graham Martin, *Member, IEEE* and Abhir Bhalerao, *Member, IEEE*

***Abstract*—Natural images contain considerable self-symmetric redundancy. In this paper, an affine symmetric model is considered for image. It provides a flexible scheme to exploit geometric redundancy. A patch of texture in a region is rotated, scaled and sheared to match other similar regions, revealing the self-symmetry relation. The general scheme for image is derived in following three steps. A texture model is devised that identifies structural patterns. Then, at a local level, the affine relation is estimated between two patches of textures, the objective being the structural patterns fit to each other. The methodology is then extended at a global level to exploit the self-symmetry of the whole image. Further, a multi-resolution framework is utilised for affine invariant texture segmentation, by which the self-symmetry of the image is exploited across space and scale. In addition, the affine symmetric image model poses an interesting approach to address practical problems such as image compression.**

***Index Terms*—image model, affine symmetry, segmentation**

I. Introduction

An image can be defined in different ways depending on the point of view. In human vision, an image is a visual observation of a perspective view of the 3D physical space projected onto a 2D plane, namely the retina of the eye and can be regarded as a composite of projected *textures* of surfaces of a scene. One can easily perceive the 3D space back from the image using various cues such as brightness, form, texture, and colour, but structural texture plays a major role [1]. Texture is an ambiguous term but includes the tactile feel, appearance or consistency of a surface, which results from the physical surface properties such as roughness or variations of reflectance differences such as the colour on a surface.

The definitions of texture[1] found in the literature [2], [3], [4] are broadly divided into two categories: *structural* and *stochastic*.

- *Structural*: Patterns that are attributed to the repetitive element or primitives arranged in a periodic manner or according to placement rules.
- *Stochastic*: Random and aperiodic patterns that may be generated by a stochastic process as opposed to a deterministic one.

Both classes[2] of texture obviously cannot be undermined by simplifying assumptions such as uniform intensities, but the former class of texture contains variations of intensities which form certain predictable and repeated patterns. A texture in the second class lacks any predictable order. The random pattern can be replicated by a stochastic process [2] and the replica should have no immediate perceptual difference to a human observer. On the other hand a small change in structural texture is easily noticed by human vision.

[1]In the context of image processing, texture is referred to as the visual appearance of the surface

[2]The texture could be further classified into more than two groups but the focus aspect of texture here is the structured-ness.

Fig. 1. Example of local affine symmetries within a natural image

A. Problem Formulation

The texture is an important visual cue to our vision. For example, being able to spot a *jaguar* by its blob patterns in a split second even against a complex background. This is a high level interpretation task in visual observation, which would take a modern computer an enormous amount of time to perform. What are the visual processes that allow one to identify a certain texture or to separate objects from the background using a texture cue? The answer to this question is still being sought. In fact, very little is known about the process [5]. From a biological perspective, the existence and role of the neurons located in the primary visual cortex indicates that our brain processes visual observation as a combination of directional bases localised in orientation and frequency [6], [7], [8], [9]. The constituent directional structure of texture pattern stimulates corresponding neurons of orientation and frequency. This is also supported by statistical research reporting that a linear combination of directional bases forms textures [9].

B. Strategy

In this work, we attempt to represent image with structural textures and their interactions by an affine transform. An overview from texture perception to the affine symmetric

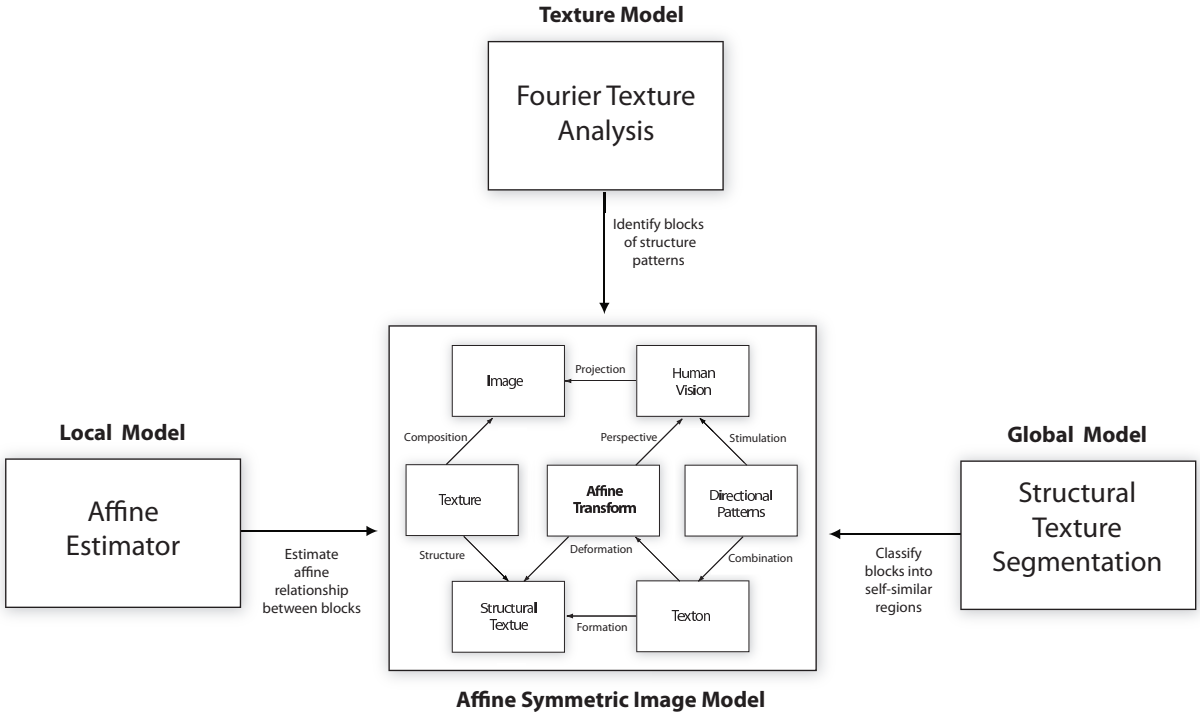

Fig. 2. Texture perception and self-similarity relations. The texton as a directional pattern descriptor and affine transforms enable the effective modelling of structural texture.

image model is illustrated in Fig. 2. A discovery of the patch-to-patch dependency would segment the image into affine symmetric regions. An example of the affine symmetry in a natural image is illustrated in Fig. 1 where a patch of hair with a simple directional pattern can be transformed to match the contour of the hat or shoulder. The idea appears simple, but it raises non-trivial problems.

- How can the structured-ness of a texture patch be determined and distinguished from non-structural patterns?
- How efficiently can an image be segmented into self-similar regions?

C. Motivation

This is interesting not only psychologically, in the sense of mimicking the highly abstract process of region identification in the brain with a simple mathematical expression, but also from the point of view of information theory as this shows potentials in reducing information redundancy and leads to a compact representation. The aim of this work is to provide a framework for affine symmetry exploitation in natural images, as illustrated in Fig. 2.

In the following, Section II introduces the affine symmetric image model and Section III describes texture spectral modelling and affine invariance texture signature. Application of the affine symmetric image model to texture segmentation task is given in Section IV. Finally, conclusions are drawn in Section V.

II. Affine Symmetric Image Model

The model that we adopted is based on texton theory [10], [11] and directional pattern recognition. Let a pattern f which is localised in frequency ξ, orientation θ, and location $\mathbf{u}$ be the most primitive unit in texture as the same type of function as shown to exist in visual perception [12], [8], [6], [13]. A linear combination of such signals forms a micro structure of texture often referred to as a *texton*. An analogous assumption is noted in [11], that

$$\textit{texton} = \int_{\xi}\int_{\theta} f_{\xi,\theta}, \qquad \textit{texture} = \int_{\mathbf{u}} T(\textit{texton}_{\mathbf{u}}),$$

Texture is defined as a spatial distribution of affine transformed textons (obtained by transform T), superimposed on a uniform lattice. It is the process of constructing texture that forms the focus of this work. Application of this model is not limited to a class of image with homogeneous content[3] but can be applied to natural images in which a number of textons are needed to represent edges and various forms of primitive patterns.

The process T of transforming a texton to another form in a geometrical sense is defined by an affine transform, a map that connects two vector spaces. The standard affine transformation, T in $\mathbb{R}^2$ space is defined as

$$T\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A_{xx} & A_{xy} \\ A_{yx} & A_{yy} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}. \qquad (1)$$

It consists of a linear transformation A followed by a translation t, where the linear transform may be composed of one or several of the following; *rotation*, *scaling*, *shear*. A series of affine transforms can be combined into a single expression.

[3]referred to as 'texture' so far, a natural image is viewed as a class of texture composites, each texture made of textons including various forms of directional patterns.

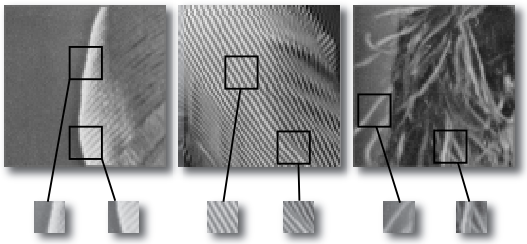

Fig. 3. Directional structure in various forms (local window size of 16×16).

The use of an affine transform to express the inter-relationship of textons is attributed to the way that images are perceived. Human vision takes visual data projected on the retina and perspective vision creates the affine symmetry of the image; texture is distorted with regards to position when projected. The greater the angular gap between the normal direction of the actual texture surface and the observer's viewing direction, the greater that distortion caused. In general, this is not visibly obvious when the focal length is substantial, but affine symmetry still exists in a region of same texture.

III. Affine Invariant Texture Signature

Knowing *what* is present in an image patch is crucial to determining affine relation between patches. Directional information is an important component of both natural and synthetic images, to which human vision is sensitive.

A. Spectral Texture Modelling

Suppose there is a directional linear discontinuity or harmonics along the line $\theta = a$ or

$$x\cos(\theta) + y\sin(\theta) = t, \tag{2}$$

then the polar representation of its Fourier transform (FT) will exhibit rapid decay as the distance from the origin $|r|$ approaches ∞, except along the line at orientation θ, where it will decay at the rate of $1/|r|$, at best [14]. In short, the decay rate along the radial lines gives the significance of directional information at θ. A 2D Gaussian model is used to model the anisotropic shape of the magnitude spectrum [15],

$$G(x|\mu,\sigma) = \frac{1}{\sqrt{(2\pi)^2|\Sigma_i|}} exp(-\frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T), \tag{3}$$

where Σ is a covariance matrix, which determines the geometric shape of the Gaussian. The covariance Σ is determined by the strength of the directional contents using the inertia tensor of the spectrum using $\sum_{\vec{\omega}} |\hat{f}(\vec{\omega})|^2 \vec{\omega}\vec{\omega}^T$. A Gaussian model on the energy distribution is assumed along the *half* radial slice ($0 \sim \pi$) of significant direction θ, whose energy dispersion (controlled by deviation, σ) can indicate the degree of regularity of the texture pattern in orientation θ, provided that other distracting frequency components are removed. Natural patterns in local windows are often multi-directional as shown in Fig. 3, but a single directional feature assumption is limiting when modelling spectra of multi-directional energy distributions.

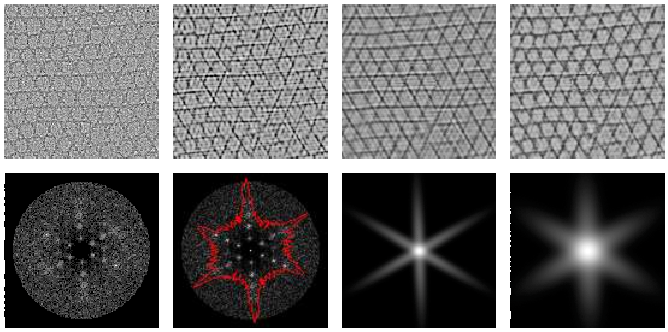

Fig. 4. Comparative evaluation of the spectral filtering on `reptile`. The polar-gaussian spectrum model achieves the best noise suppression while maintaining the structural integrity of the texture (top to bottom : reconstruction and spectral model, left to right : Noisy spectrum, Polar contour model, Gaussian mixture model and Polar Gaussian model

Let $S(\theta)$ be a normalised Fourier slice integration of the magnitudes, resulting in an adaptively shaped frequency contour that allows modelling spectra of multiple directional components without prior knowledge of the number of energy clusters.

$$S(\theta) = \frac{1}{c_\theta} \int F_P(r,\theta)\, dr, \tag{4}$$

or on a Cartesian grid,

$$S(\theta) = \sum_x \sum_y |F_C(x,y)| \delta(x\cos\theta + y\sin\theta), \tag{5}$$

where c_θ is a normalising factor determined such that the resulting contour includes more than a certain percentage of the total energy along the Fourier slice. The contour represents the energy distribution of the dominant directional pattern.

The contour can be used to distinguish the important coefficient from others by converting the contour to a binary mask $m(x,y)$ so that coefficients outside the shape are set to zero [16] (see Fig. 4).

$$\tau_{approx} = \sum_x \sum_y \frac{|F_C(x,y)|\delta(m(x,y))}{|F_C(x,y)|}. \tag{6}$$

The analysis window size is confined to $16 \leq N \leq 64$ which the contour model captures pattern well. Care must be taken in normalising the contour before thresholding so that the significant energy clusters are included inside the shape by controlling the ratio of the energy covered to the total energy τ_{approx}. Otherwise, important structure will be lost and artefacts created, for example the ghosting artefacts sometimes referred to as the *Gibbs phenomenon*, due to suppressing too many high frequency coefficients, as shown in Fig. 4. The binary decision for determining the significance of a coefficient is ill-defined. An ordering of coefficients according to significance is desirable.

A solution to mitigate the Gibbs phenomenon as well as to construct efficiently a multi-directional window is to employ a Gaussian window in a slice-wise manner. A 1D Gaussian window per Fourier slice, centred on the DC coefficient (due to the Hermitian symmetry) is given as

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \tag{7}$$

where σ denotes the deviation of the slice, which can be computed simply as follows:

$$\sigma = \sqrt{\frac{2}{N}\sum_{i}^{N/2}(i \cdot x_i)^2}, \tag{8}$$

where i is the index and N denotes the number of samples on the Fourier slice. The choice of using the Gaussian function on the slice is due to the uncertainty principle, that is the Gaussian function achieves optimal spread in space and frequency. It is also smooth in both domains due to the modulation theorem. All the resulting 1D Gaussian windows are normalised such that $G(0) = 1$, and then the windows are modulated in proportion to $\omega(\theta)$ in Eq.(9).

$$\omega_\theta = \frac{\sum_r |F_C(x,y)|\delta(x\cos\theta + y\sin\theta)G_\theta(r)}{\sum_\theta\sum_r |F_C(x,y)|\delta(x\cos\theta + y\sin\theta)G_\theta(r)} \tag{9}$$

The Gaussian window in the slice representation is converted to the Cartesian form using Eq.(10).

$$x = r\cos(\theta) \qquad y = r\sin(\theta) \tag{10}$$

The model provides a multi-directional filter, which allows slice-wise treatment and, importantly, it can be obtained without any prior knowledge or nonlinear estimation technique, (*cf.* Gaussian mixture modelling). A comparative illustration of spectral filtering using different spectral models is shown in Fig.4. The polar Gaussian model is shown to fit the energy spectrum best. Preliminary results on denoising were presented by the authors in [16]. A similar concept of spectral modelling were reported recently in [17].

B. From Global to Local Analysis

Having introduced a local texture model for the recognition of directional patterns, a useful step forward is to exploit the global relationship of local blocks across scales to overcome the problem of a fixed window size. Locality can be introduced to the FT by windowing the image around the position of interest. When using an arbitrary window function this transform is the Windowed Fourier Transform (WFT), which generalises the more familiar Short Time Fourier Transform (STFT) for signal analysis. In [18], Wilson et al. introduced a multiresolution version of the STFT which generates an overcomplete wavelet decomposition of the image using appropriately windowed Fourier bases functions. The MFT introduces scale to the WFT by varying the size of the window according to a scale parameter. With the windowing function $g(t)$, the transform of a function $f \in L^2(R)$ at position $\mathbf{u}$, frequency ξ and scale s is defined as follows:

$$Mf(\mathbf{u},\xi,s) = \frac{1}{s}\int_{-\infty}^{+\infty} f(t)w(\frac{t-\mathbf{u}}{s})e^{-i\xi t}dt. \tag{11}$$

The MFT contains redundancy as the whole frequency content of the image is represented at each scale. The redundancy could be a drawback in compression applications but in the image analysis domain it often acts as a correction term that can be used to rectify a decision by cross-scale analysis. An appropriate sampling strategy such as the use of an image pyramid can reduce the redundancy significantly.

The discrete MFT, as described in [18], is derived by sampling all three parameters at intervals, determined by level. *Level* refers to the evaluation of the MFT at a single scale. The discrete MFT consists of a number of levels, whose corresponding scales are the sample points of the scale variable. For a given level and spatial sample point, the frequency samples form the local spectrum corresponding to the region of the image that the spatial window is concentrated on. Let $f(\mathbf{u})$ be a discrete image of size $2^M \times 2^M$ with $M \in \mathbb{Z}^+$, l be level and the spatial sampling between adjacent levels differs by a factor of 2, then the total number of levels is given by $\log_2 2^M (= M)$. The number of spatial samples at level l is given by 2^{M-l} and the number of frequency samples at level l is given by 2^l when critically sampled and 2^{l+k} when oversampled (Fig. 7). A window function, $w(y)$ is used that has good joint localisation in space and frequency and is applied with a 50% overlap across the image, which enables artefact free synthesis of the spatial domain to the spatial domain. An appropriate window function is the squared cosine:

$$w(\vec{y}) = \cos^2[\pi p/2N]\cos^2[\pi q/2N], \tag{12}$$

where $\vec{y} = (p,q)^T, 0 \le p,q \le N$.

C. Affine Invariant Signature

Spectral signatures such as the annular-ring and wedge samples have been used as features in various applications for texture discrimination purposes but they are not invariant to geometric deformation. A affine invariant texture description is presented.

The texture signature is based on a contour of the spectral texture model in section III-A. The contour model C of the multi-directional spectrum is given by points traversing along the contour of the polar Gaussian,

$$C = \{(r,\theta)\,|\,r = S(\theta)\}. \tag{13}$$

Given a contour signature $C(x,y)$, its constituent boundary pixels are traversed to yield a parametric equation based on the affine length of a closed curve, as shown below. This is linear under affine transformation and also yields the same parameters, independent of the initial representation.

$$\int_C \sqrt[3]{\dot{x}(t)\ddot{y}(t) - \dot{y}(t)\ddot{x}(t)}dt, \tag{14}$$

where the number of dots indicates the order of the derivatives. Having encoded the boundary as a function of the parameter, taking the Fourier transform of the boundary equation results in $[U,V]^T$, where U and V are Fourier coefficients referring to the x and y coordinates respectively. Since the Fourier transform is a linear operator, the equation below holds,

$$[U_k, V_k]^T = A[U_k^0, V_k^0]^T, \tag{15}$$

where $[U_0, V_0]^T$ denotes the same coefficients from the affine transform of the reference block. By including another pair of coefficients and extending Eq.(15) to a 2×2 matrix, obtaining the determinants reveals a linear factor. A simple division of both sides by one side produces an absolute affine invariant feature. For more details, the reader is referred

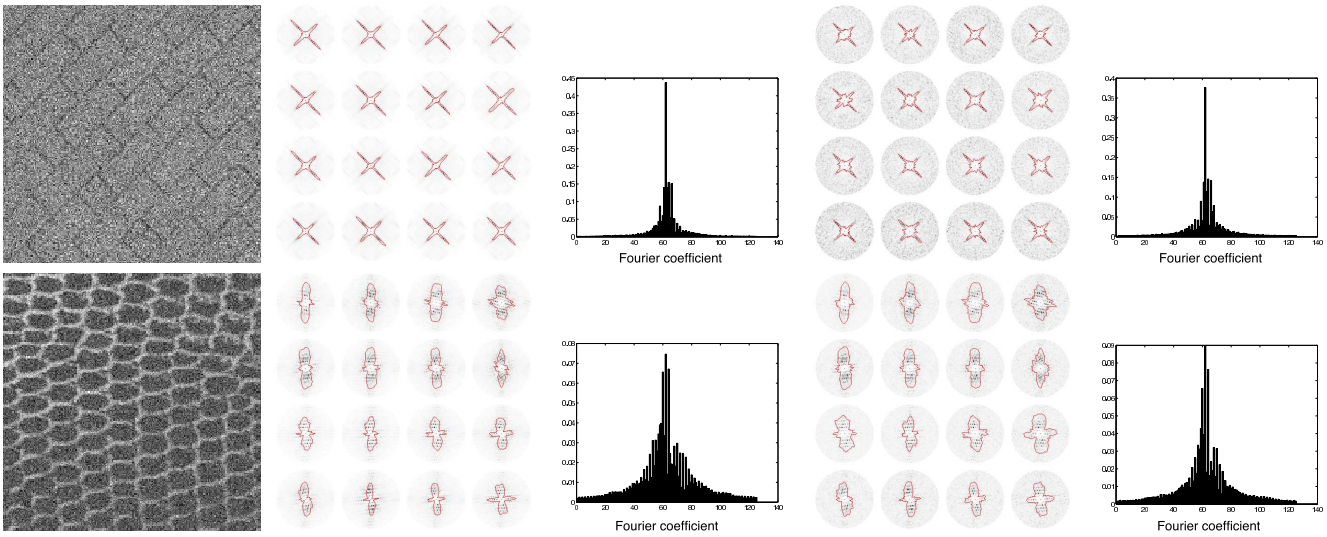

Fig. 5. Noise robustness test for `polar contour` (from left to right : source, polar contour, Fourier description, polar contour at 15 dB and Fourier description at 15 dB)

to [19].

Fig. 5 shows a simple noise robustness test of the polar signature extraction, where the sample textures are of size 128×128 pixels. The first column shows two test textures, to which white Gaussian noise is added to give a SNR of 15dB. The sample is partitioned into 64×64 blocks with 50% overlap and the polar contour is extracted in each block, resulting in 4×4 contours. The middle column shows the polar contour extracted from the original texture and the last column shows the contour extracted from the noisy textures in the first column. As shown, shapes of the extracted contours (red) and distribution of the Fourier description coefficients are consistent regardless of the presence of noise. It should be noted that different shapes of polar contours and Fourier descriptions depend on the textures.

Fig. 6 shows a simple classification experiment, where blocks are 64×64 pixels and 50% overlapped. The composited texture in the first row consists of three kinds of texture, whose number of significant directional features are different to each other. The extracted contour signatures are shown in the middle column and their Fourier description is visualised in the scatter graph. It is important to bear in mind that the fundamental assumptions underlying the affine-invariant descriptors have been violated. The observed textures are strictly not affine transforms of some prototype. Had this been true, the distributions of both scatter graphs would be reduced to single points. The second texture in the new row consists of two classes of texture, a pair of which is a rotated version of the other. Both textures have the same number of directional features but the statistical variation along the contour is different. This can be visualised in the corresponding scatter graph, where the distribution of the features comprises a line.

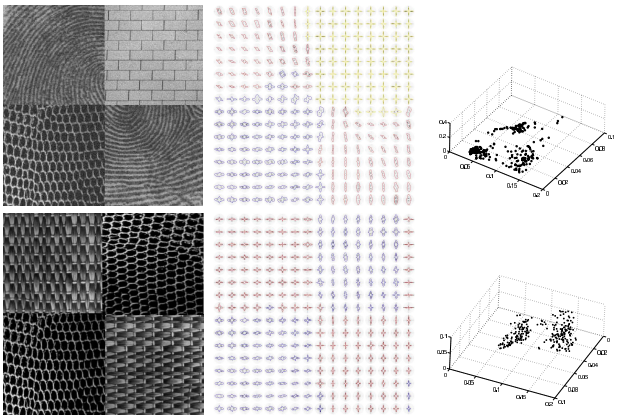

Fig. 6. Segmentation test of texture `composites`: texture composite (512×512) and classification result in pairs, white Gaussian noise is added to both texture samples (15dB in SNR)

IV. Application to Texture Segmentation

Segmentation has been studied for decades and utilised in many applications requiring the detection, recognition and measurement of objects in images. The aim is to model an image utilising affine symmetry, which in a local manner, is the defining of a geometric relationship between local blocks [20]. Discovering the optimal groups that minimise the overall warping error is non-trivial, and known methods are computationally intensive. The problem is approached by extracting affine invariant features from the local blocks (not requiring affine computation). Considering that texture exhibiting strong 'directional pattern', the Fourier representation offers an ideal way to extract features as follows.

- The Fourier power spectrum is invariant to translation and only the linear part of the affine transform needs analysis.
- Fourier slice analysis allows the efficient extraction of both directional patterns and regularity of content.
- The Fourier description of the polar signature is simple

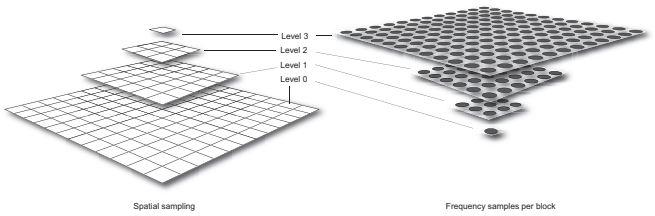

Fig. 7. Multiresolution Fourier Transform built on the Laplacian pyramid

and fully affine invariant.

A. Segmentation by Searching and Grouping Affine Symmetric Regions

Let f_i be a subblock from a uniformly partitioned image f and B be a pool of subblocks. The objective is to sort subblocks into a desired number of affine invariant groups. This requires identification of the optimal prototypical (centroid) block of each group that minimises the overall transformation error for all combinations of prototype and other target blocks. The brute-force search leads to the discovery of the optimal prototype block as well as affine invariant segmentation (grouping), but it is clearly a computationally arduous task given the number of possible combinations. The computational burden can be halved by assuming that $\forall F_i, F_j \in B \quad |T(f_i, f_j) - T(f_j, f_i)| = 0$,, where $T(f_i, f_j)$ transforms f_i to fit f_j. Nonetheless, the computational requirement becomes prohibitive as the image size increases, with $O(n^2)$ for the affine transform where n is the number of local blocks.

Some warping based methods exist, and they are recapped here. Wilson and Li [21] performed texture segmentation using affine symmetry. They used the warping error as one of the distance metrics in a Multiresolution Markov random field framework. The method requires relatively less computation than the exhaustive search as the affine transformations between the current block and its neighbouring patches are computed. Later, Bhalerao and Wilson [22] developed an algorithm based on the translation invariant property of the Fourier magnitude spectrum. It reduces the computation by using the Fourier magnitude spectrum as a single long feature vector, and gains invariance by having an affine symmetric group of vectors as the centroid, the members of which are derived from a single block by scale and orientation changes. Smith [23] utilised a metric to measure the degree of deformation from an affine matrix and the transitivity of the matrix. Affine transforms between blocks and one prototype is computed, then using transitivity, blocks are classified such that the overall deformation is minimised. This is based on an assumption that only a small amount of deformation exists between the blocks in a self-similar group, which may not be true in natural images. The latter two methods in particular require human assistance to select an adequate number of prototypes.

level=2 *level=1* *level=0*

Fig. 8. Polar contour pyramid : `lena`.

B. Unsupervised Segmentation by Independent Component Analysis

The main purpose of feature extraction is to map differences in spatial structures, either stochastic or geometric, into difference values in higher or lower dimensional feature spaces, where one group of data points can be distinguished from another in a semantic sense. Most segmentation methods follow the same general strategy.

An unsupervised block classification based on the number of directional features using Independent Component Analysis (ICA) is attempted. ICA and its variants provide representations that utilise a set of linear basis functions [9], [13]. What the approaches have in common is that they try to reduce the information redundancy by capturing the statistical structure in the images, beyond second order information. ICA finds a linear non-orthogonal coordinate system (basis) in multivariate data, determined by the higher-order statistics, which is well localised in frequency and orientation. This experiment is motivated by the assumption that ICA can allow structures between blocks to be found, as well as how similar they are to each other.

Despite the fact that ICA identifies common directional linear structures, it is not obvious how to extract affine invariance between blocks with only the discovered basis functions. However, attention is directed towards the number of directional linear bases in each block captured by ICA, which appears affine invariant according to the trials illustrated in Fig. 10. The number of bases with large coefficients (active) in a block seems dependent on the complexity of the texture, considering ICA bases are well-localised in orientation. With reference to figure 10, the crack in the log, composed of only one basis at a specific orientation, is distinguished from other textures while the jaguar's blob texture is not captured as it requires a larger analysis window. This makes sense as a block is unlikely to fit one to another if the number of directional linear features is different. The experiment is conducted as follows:

1) The image is decomposed using the Laplacian pyramid
2) A subband ($level = 0$) is tiled by 16×16 pixel blocks and the local blocks are fed into ICA.
3) The different number of significant basis functions for each block can be found as

$$\text{argmax}_k \left\{ \frac{\sum_{i \in s}^{k} w_i}{\sum_i w_i} \leq c \right\} \qquad 0 \leq k \leq k_{max}, \quad (16)$$

where w_i is the weight of the basis function b_i, c is the percentage (set at 0.9) of energy that a set of significant basis coefficients in the total energy of the block occupies, and s is a set of basis function indices sorted by $w_i / \parallel b_i \parallel$. k_{max} is the maximum number of basis functions for a block.

4) The blocks are grouped based on the number of significant basis functions in each block. Blocks in each group should be compared for better classification if they are not affine invariant. This can prevent different texture blocks with the same number of directional features from being in the same group. However, this experiment is to discover the effectiveness of the number of directional features as an affine invariant feature.
5) In [24], for each group, a prototypical block is given by

$$\text{argmax}_{k \in K} \left\{ \sum_i \frac{|w_i^k|}{e_2^k} \right\}, \tag{17}$$

where e_2^k is the eigenvalue of the second principal component of block k.

This results in a block composed of strong directional bases with high weights. This is to verify that the classification result does conform with the purely warping error-based classification. Despite the brute-force search not being conducted, the quality of the reconstruction indicates a good fit between prototype and other blocks in the class. In the case of the undecimated Laplacian pyramid, the block size should be increased by a factor of 2 for multiresolution analysis.

The ICA based method has a couple of problems. Firstly, the computational burden of ICA is significant. Secondly, it does not exploit inter-scale correlation, which is considered an important factor to tackle not only the fixed size of the local window but also to avoid any bias resulting from a single scale. Nonetheless, performing ICA on the high level of the pyramid is difficult due to an insufficient number of observations (blocks), as the number of local blocks decreases by a factor of 4 as the level increases.

C. Cluster based Segmentation

With regard to the window size for texture analysis, there is a problem known as the class-boundary uncertainty, *i.e.* if the analysis is confined to a small window, a better resolution of segmentation is obtained but confidence of the texture characteristics within the window is lost. On the other hand, a larger window allows a better analysis of texture but results in a coarse resolution of segmentation. A solution is to employ a multi-resolution approach. Texture information at the top level passes down to the lower level and is combined recursively as it proceeds. The entire image is covered initially by a single window. The window is divided into four small uniform windows in accordance with a quadtree structure where the analysis information is passed on to the next level. The procedure iterates until all regions are uniform or until the desired number of regions have been established.

Having applied the MFT [18] to a source image, the affine invariant features are extracted from the Fourier spectrum at each scale. The resultant features, v, at the next level are joined together with the quadtree parent as follows:

$$Feature(i,j) = \{p^k(V) | 0 \leq k \leq l\}$$
$$p^k(V_{i,j}) = V_{\lfloor i/2^k \rfloor, \lfloor j/2^k \rfloor, k} \ , V_{i,j,k} = w_k \times \{v_{i,j,k}\}, \tag{18}$$

where w_k and k refer to a weight and level respectively. The combined features are fed into a clustering algorithm. Expectation Maximisation (EM) with a Gaussian mixture is used for clustering. Prior knowledge of the number of classes is required due to the nature of the algorithm. The EM clustering is chosen only for simplicity as the main focus is to determine an effective affine invariant feature. A random field approach may be of interest if convergence of the number of classes is desired [21], [25].

The MFT used here is implemented slightly differently from that in [18]. The local block size is adjusted to change the size of window according to *level* while keeping the resolution fixed. This is to use only a specific high frequency band, which is assumed to include pure texture information. The applied settings are as follows:

- The decomposition level is $J = 2$. The high frequency subband is extracted using the Laplacian bandpass filtering and the MFT is computed. The MFT of each image is given in Fig. 8.
- The local blocksize is set to $2^{4+l} \times 2^{4+l}$ where l is *level*.
- The number of angular segments, Λ, for Fourier slice analysis is set to the width of the block. The polar contour is parameterised at a uniform angular interval, π/width.
- Affine invariant features include all the Fourier coefficients except DC and two adjacent coefficients.
- w_k in Eq.(18) is set to $1/2^l$ where l denotes *level*.

Firstly, it is observed that the signature extracted from the Fourier transform is too jagged in severe levels of noise (e.g. $0dB$ in SNR) due to the scattered high-frequency coefficients, which in turn disrupt the affine invariant description and results in a poor clustering. The scattered high-frequency coefficients are removed using universal thresholding [26], Eq.(19)

$$\tau_s = \alpha \frac{\sqrt{2 \log 256} \sigma}{1.23^{J-s}}, \tag{19}$$

before the Fourier slice projection and a Gaussian smoothing filter is applied to the shape boundary. This, in fact, makes the shape-extraction robust to noise [16]. The polar contour in a small block is still affected by noise, but it is alleviated as the block size increases.

Secondly, many contour signatures from the bottom level of the MFT are elliptical, which makes affine invariant shape description useless. Starting with a window of 32×32 at the bottom level still produces an acceptable result. However, it is found that using the $area$ of the contour at the bottom level feature produces a better result. This is because the shape size increases with the strength and directionality of the feature and decreases as the directional pattern is less significant. It is probable that different textures with the same size of polar contour fall into the same class, but it is the most useful information that is collected at the bottom level. The discrimination gets better as the structural information from a

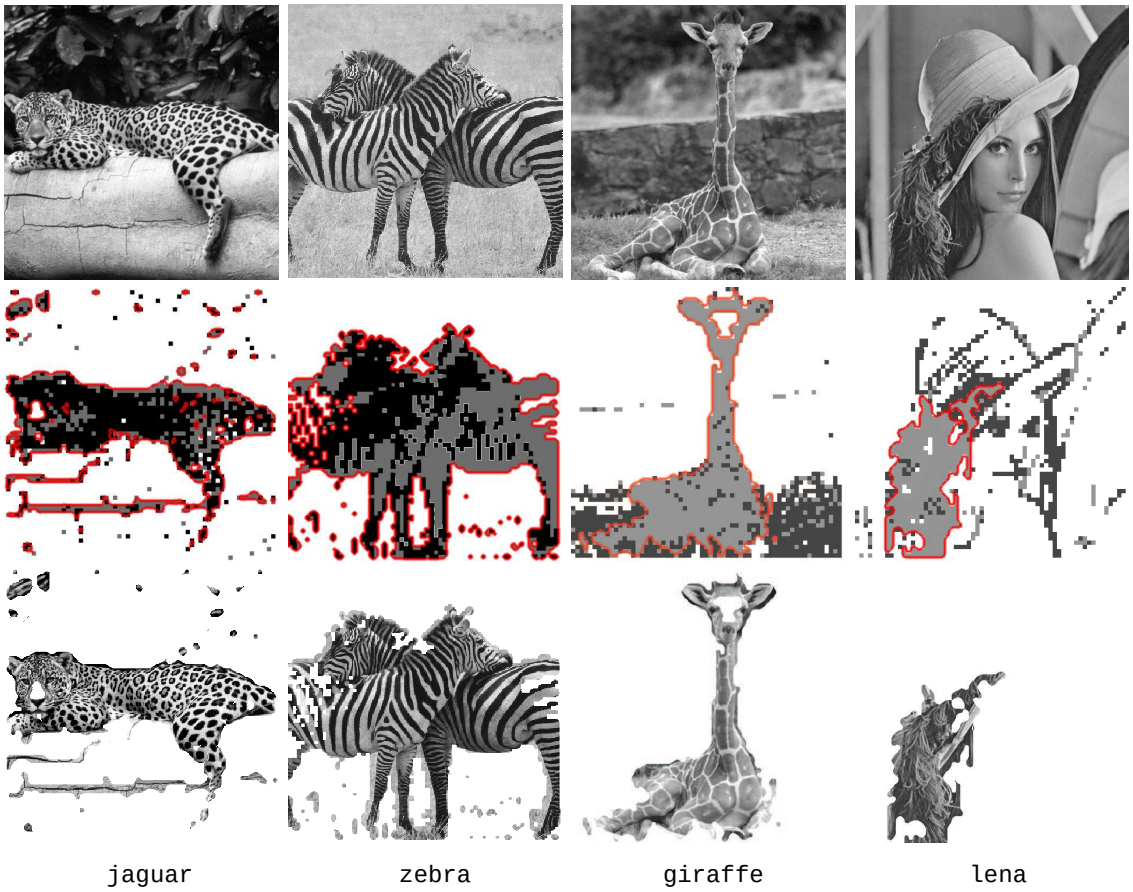

Fig. 9. Block classification test on natural images

bigger window is reflected; passed on from the parent block and combined together by Eq.(18)

Lastly, the block size of 16×16 results in a coarse segmentation map, the effective resolution of which is 8×8 pixels per block considering the 50% overlap. It is assumed that the class with the most complex texture based on Eq.(20) (blocks) is the region of interest,

$$R(s) = \frac{\oint s}{\oint CH(s)}, \tag{20}$$

where CH is a convex hull function that returns results in the form of a set of coordinates. A morphological operation is applied to the binary map of the assumed region of interest using a 6×6 disc structuring element to obtain a smooth contour of the segmentation boundary and remove possible outliers.

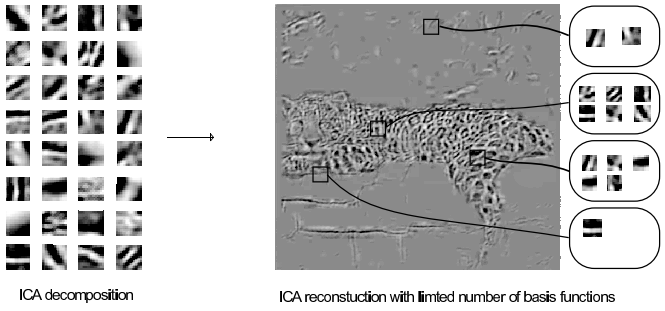

Fig. 10. ICA based block classification: a coarse reconstruction formed from a limited number of ICA bases and a different number of bases for different blocks

D. Experimental Evaluation

Two preliminary experiments were presented earlier to illustrate the effectiveness of the new feature. Fig. 5 shows the contour signature extraction in the presence of noise. Fig. 6 tests segmentation of simple composites of structural textures to which Gaussian noise is added to give a SNR of 15dB.

With the successful test on the sample image test set, the method is applied to four natural images - `jaguar`, `zebra`, `giraffe`, and `lena` as shown in Fig. 9. The segmentation results are shown in pairs - the classification map with a red contour obtained by the morphological operation, and a segmented image obtained by clipping out the region inside the red contour. Three kinds of textures are assumed in all the test images (three component Gaussian mixture model). `jaguar` is segmented by blob, crack in the log and the rest. The blob and crack differ in the number of directional features and it is well captured in the classification. `zebra` is segmented by stripe, grass and the remaining region. The stripe texture is diverse in scale but it is captured as a single region. `giraffe`

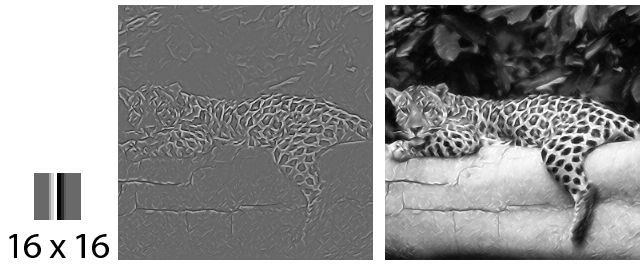

Fig. 11. Image approximation with a randomly chosen prototypes : `jaguar`; the approximation is shown as a synthesized highpass and a reconstructed image

is segmented by the giraffe skin, grass and the remaining region. The three regions are separated very well with only few outliers. `lena` is segmented by the fur on the hat, the edge and the remaining region. The first two textures differ largely by the number of directional features but blocks in the fur region are of a single direction feature, with some periodic and others of linear discontinuity or lines. The two regions are well separated. The classification result on all test images has many outliers. Although this may be improved by the employment of a sophisticated algorithm, the result is very promising considering the textures of the natural images are not exactly affine symmetric despite the fact that they look very self-similar.

V. Summary and Discussions

We have described an affine symmetric image model that defines patch-to-patch affine relationships on an image with a uniform lattice. Considering textures on every patch, it was realised that not all patches are appropriate for the exploitation of affine symmetry with other patches, particularly patches from background areas with uniform intensity. The first obstacle, therefore, was to develop a texture analyser so that a patch with structural texture can be distinguished and treated differently from non-structural textures. Two different approaches to estimate the distance in affine space were presented; one based on warping residue and the other based on affine invariant features. The latter provides a more practical solution in terms of computational efficiency. Affine invariance has received much attention with the recent emergence of content based retrieval systems, which could take advantage of this work. The majority of existing texture analysis methods, however, are not designed to analyse texture from an invariance viewpoint. Several noteworthy geometric invariant analysis methods have the same common theme of directional pattern recognition. This led us to develop further the texture model used for structured-ness analysis into a fully affine invariant descriptor. The usefulness of the new affine invariant feature was demonstrated in a multiresolution framework for the segmentation of a textured object. This work not only presents an interesting approach to the segmentation task but also offers a feasible solution for efficient implementation.

The underlying concept has been applied to image classification by many researchers but few have applied the affine symmetry model to segmentation by partitioning the image into blocks. The complexity of the algorithms, however, has been a major issue prohibiting practical implementations. The motivation has been to develop a computationally efficient image texture classification algorithm while maintaining the texture discriminative power of previous approaches. The simplicity and efficiency of the presented approach utilising an affine invariant shape description is demonstrated. It may be of interest where efficient texture segmentation is required. Experimental evaluation indicates acceptable segmentation results for structural texture and the algorithm's robustness to noise. Further study utilising a random field segmentation framework with other useful features may improve the algorithm, thereby determining the optimal number of segmented regions. Additionally, it can also be utilised for image compression.

A. Application to Image Approximation

Considerable reported research has been directed towards reproducing an image from a compact representation using statistical properties. Despite the wide variety and large complexity of many data compression techniques, they are all based on the same simple principle. This requires a statistical characterisation of the data to be compressed, with which a measure of probability or frequency of occurrence is associated, thus achieving compression. In contrast, very little progress has been made towards image compression from understanding the image content. The approach would allow compression as well as an analysis of the image content and the incorporation of new functionality such as the identification of an object by texture. Some results of our initial experiments are shown in Fig. 11 where a single texture patch is used to approximate the entire image. A combined approach with PCA is introduced in detail in [27]. A better solution, however, would determine the number of textures present in the image and use a representative patch from each texture region to approximate the image region by region. This is ongoing research built on our segmentation algorithm and we are currently developing an efficient affine parameter entropy coder. Preliminary results were presented by the authors in [24].

References

[1] J. Gibson, *The perception of the visual world*. Houghton Mifflin, 1950.

[2] T. Hsu and R. Wilson, "A two-component model of texture for analysis and synthesis," *IEEE Trans. on Image Process.*, vol. 7, no. 10, pp. 1466–1476, October 1998.

[3] R. Haralick, "Statistical and structural approaches to texture," *IEEE Proc.*, vol. 67, no. 5, pp. 786–804, May 1979.

[4] R. Haralick, K. Shanmugam, and I. Dinstein, "Texture features for image classification," *IEEE Trans. SMC*, vol. 3, no. 6, pp. 610–621, November 1973.

[5] B. Olshausen and D. Field, "What is the other 85% of V1 doing?" *23 Problems in Systems Neuroscience*, 2004.

[6] ——, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 1996.

[7] E. Oja, "A simplified neuron model as a principal component analyzer," *J. of Mathematical Biology*, vol. 15, pp. 267–273, 1982.

[8] R. P. N. Rao, B. A. Olshausen, and M. S. Lewicki, *Probabilistic Models of the Brain: Perception and Neural Function*. MIT Press, 2002.

[9] E. O. A. Hyvärinen, J. Karhunen, *Independent Component Analysis*. Wiley and Sons, 2001.

[10] B. Julesz, "Texture gradients: The texton theory revisited," *Spatial Vision*, vol. 1, no. 1, pp. 19–30, 1985.

[11] S. Zhu, Y. Z. W. C. Guo, and Z. J. Xu, "What are textons?" *IJCV. Special Issue on Texture*, 2005.

[12] B. Olshausen and D. Field, "Sparse coding with an over-complete basis set: A strategy employed by V1?" *Vision Research*, vol. 37, pp. 3311–3325, 1997.

[13] B. Olshausen, "Learning sparse, overcomplete representations of time-varying natural images," in *IEEE ICIP.*, 2003, pp. 41–44.

[14] E. Brigham, *The fast Fourier transform and its applications*. Prentice Hall, 1988.

[15] Z. Yao, N. Rajpoot, and R. Wilson, "Directional wavelet with Fourier-type bases for image processing," in *Wavelet Analysis and Applications*, ser. Applied and Numerical Harmonic Analysis, X. Y. T. Qian, M.I. Vai, Ed. Springer-Verlag, 2007, pp. 123–142.

[16] H. Park, G. Martin, and Z. Yao, "Image denoising with directional bases," in *IEEE ICIP.*, San Antonio, Texas, USA, Sep. 2007, pp. 301–304.

[17] D. Hammond and E. Simoncelli, "Image modeling and denoising with orientation-adapted gaussian scale mixtures," *IEEE Trans. Image Processing*, vol. 17, no. 11, pp. 2089–2101, Nov 2008.

[18] R. Wilson, A. Calway, and E. Pearson, "A generalized wavelet transform for Fourier analysis: the Multiresolution Fourier Transform and its application to image and audiosignal analysis," *IEEE Trans. on Image Process.*, vol. 38, pp. 674–690, Mar. 1992.

[19] K. Arbter, W. Snyder, H. Burkhardt, and G. Hirzinger, "Application of affine-invariant Fourier descriptors to recognition of 3-d objects," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 12, no. 7, pp. 640–647, 1990.

[20] H. Park, G. Martin, and A. Bhalerao, "Structural texture segmentation using affine symmetry," in *IEEE ICIP.*, San Antonio, Texas, USA, Sep. 2007, pp. 49–52.

[21] R. Wilson and C. Li, "A class of discrete multiresolution random fields and its application to image segmentation," *IEEE Trans. PAMI*, vol. 25, no. 1, pp. 42–56, 2003.

[22] A. Bhalerao and R. Wilson, "Affine invariant image segmentation," in *BMVC.*, Kingston University, UK, 2004.

[23] T. Smith, "Texture modelling and synthesis in two and three dimensions," Master's thesis, University of Warwick, Department of Computer Science, Signal & Image Processing Group, 2004.

[24] H. Park, A. Bhalerao, G. Martin, and A. Yu, "An affine symmetric approach to natural image compression," in *Second International Mobile Multimedia Communications Conference*, Alghero, Italy, Sep. 2006.

[25] C. Li, "Multiresolution image segmentation integrating Gibbs sampler ang region mergion algorithm," *Signal Processing*, vol. 83, pp. 67–78, 2003.

[26] D. D. an dI.M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[27] A. Bhalerao and R. Wilson, "Warplet: An image-dependent wavelet representation," in *IEEE ICIP.*, 2005, pp. 490–493.