THE UNIVERSITY OF
WARWICK

Original citation:
Park, Heechan, Martin, Graham R. and Bhalerao, Abhir. (2010) Local affine image matching and synthesis based on structural patterns. IEEE Transactions on Image Processing, Volume 19 (Number 8). pp. 1968-1977. ISSN 1057-7149

Permanent WRAP url:
http://wrap.warwick.ac.uk/5142

warwick**publications**wrap
highlight your research

http://wrap.warwick.ac.uk

Local Affine Image Matching and Synthesis based on Structural Patterns

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

Abstract

A general purpose block-to-block affine transformation estimator is described. The estimator is based on Fourier slice analysis and Fourier spectral alignment. It shows encouraging performance in terms of both *speed* and *accuracy* compared to existing methods. The key elements of its success are attributed to the ability to: 1) locate an arbitrary number of affine invariant points in the spectrum that latch onto significant structural features; 2) refine each estimated invariant point by taking the phase-gradient into account; and 3) directly compute all four linear parameters of the affine transform from the spectral alignment. Experimental results using a wide range of textures are presented. Potential applications include affine invariant image segmentation, registration, affine symmetric image coding, and motion analysis.

Index Terms

affine estimation, structural texture

I. INTRODUCTION

An ability to efficiently estimate geometric transformations is desired in many vision related applications. Determining the transformation parameters, which map pixels from one image to the corresponding pixels in related images, enables the comparison of images obtained from different views or time frames. The parameters could be utilised to correct misalignment [1], to determine properties such as motion, depth or shape [2], [3], or as a distance feature between pairs of images in a classification task [4]. There are various approaches to estimate the transform [5], [4], [6] however this paper seeks to explore a particular texture analysis based technique, which is termed the *affine estimator*.

A. Problem Formulation

Geometric deformation can be simplified by using knowledge of the imaging process to constrain the class of transformation being estimated. The simplest example is to assume that the deformation is only translational. This is an adequate assumption only when certain imaging conditions are maintained such that the viewpoint is perpendicular to the object and the distance to the object is constant. A more flexible approach is to assume that the images are related by a six parameter affine transformation, corresponding to scaling, rotation, shear, and translation. The standard affine transformation, T in $\mathbb{R}^2$ space is defined as

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

where A and t represent a linear part and a translational part of the transformation respectively. The question is, given two similar images (deformed under affine transformation), is it possible to accurately and efficiently estimate the affine transform?
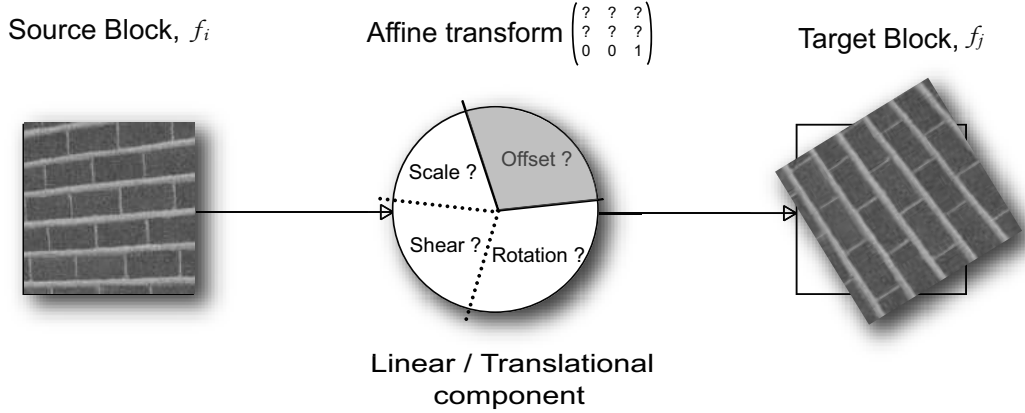
Fig. 1. Block to Block Affine transformation

B. Strategy

A straightforward approach to determining the optimal affine parameters involves a brute-force search of a six-dimensional space. It requires computation of the matrix multiplication using Eq.(1) and the corresponding error, d_{ij} shown in Eq.(2) resulting from not only every translational change but also by all possible linear deformations. The affine estimation is illustrated in Fig.1.

$$d_{ij} = |f_j(u) - T_{ij}(f_i(u))| \tag{2}$$

where $f_i(u)$ and $f_j(u)$ are the two blocks. A better approach is to regard the problem as a class of parametric optimisations that utilises affine invariant features (or at least are based on the same principle) to fine-tune the transformation parameters and minimise the warping error with minimal computation.

Hsu and Wilson's method [4] and its variants [7], [6] show merit. The method performs Fourier spectral alignment; a direct affine estimation based on the equivariance of the Fourier spectrum (without search). We develop the algorithm further by incorporating Fourier slice analysis and Fourier spectral alignment, the key motivation being the ability to:

- Locate and determine an arbitrary number of affine invariant points in the Fourier spectrum that latch on to significant structural features
- Directly compute all four linear parameters of the affine transform by spectral alignment, as opposed to the limited estimation possible with other Fourier based methods that exclude *shear*.
- Refine the estimated invariant point, by taking the phase-gradient information into account, as opposed to some methods that deal with only the strength of the directional pattern.

II. FOURIER BASED ESTIMATION METHODS: REVIEW

For the past few decades researchers have studied affine estimation for motion estimation and image registration and various methods have been suggested. The Fourier transform is often adopted due to a useful linear property. Given the Fourier transform (F_i) of the source block (f_i), and a linear relation (A_{ij}) with the Fourier transform (F_j) of a target block (f_j), a linear relation ($(A_{ij}^{-1})^T$) holds in the spatial domain too. This is the *affine theorem*.

$$\begin{aligned} F_j(\mathbf{u}) &= A_{ij} F_i(\mathbf{u}) \\ &= \frac{1}{|detA|} e^{i\mathbf{u}^T \mathbf{t}_{ij}} |F((A_{ij}^{-1})^T f_i(\mathbf{u}))| \end{aligned} \tag{3}$$

In addition to the linearity property, the Fourier spectrum is a favourite domain in which to estimate geometric transformations, since the Fourier transform decomposes the geometric transform, T_{ij} into two
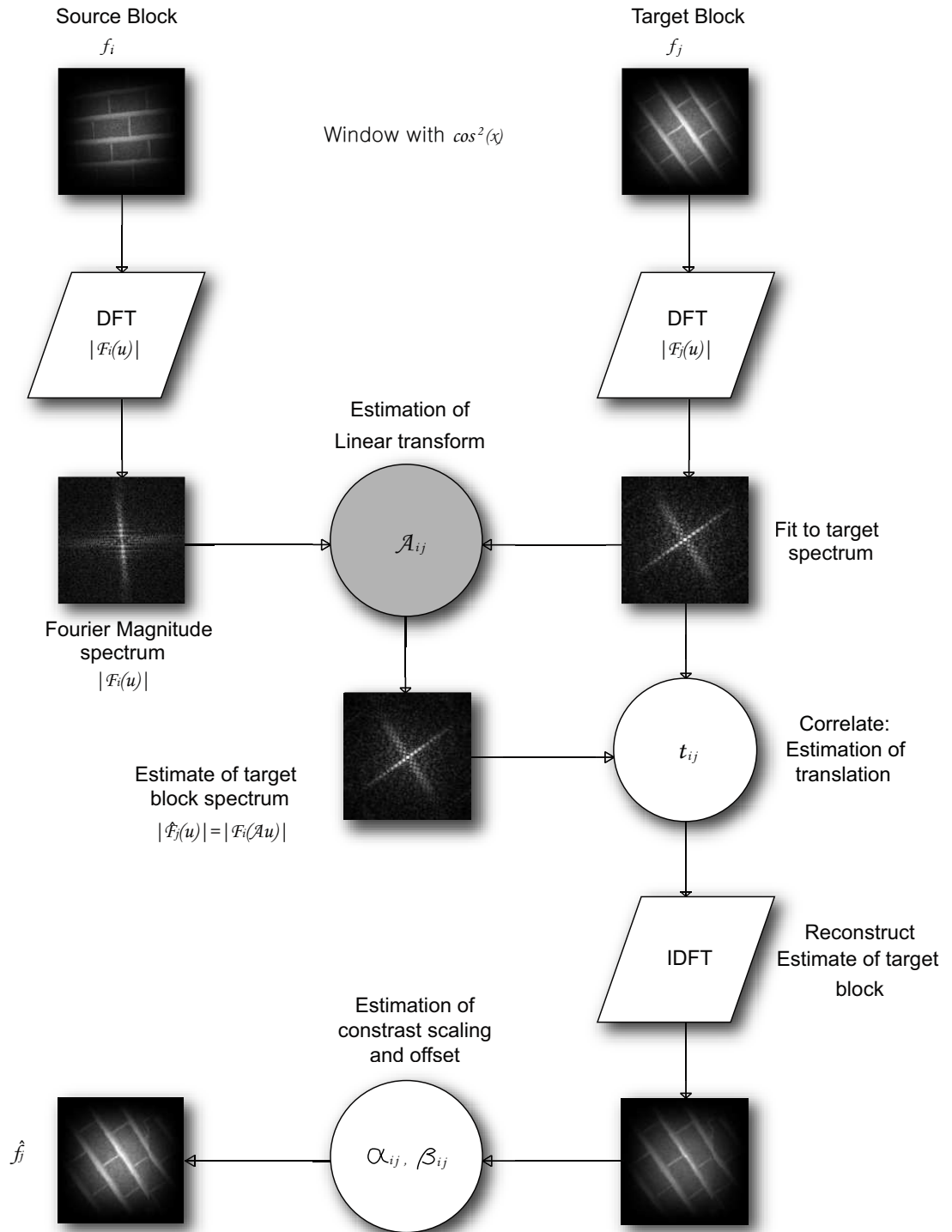
Fig. 2.   Overview of Affine transform estimation process

parts: a linear part, A_{ij} that affects only the power spectrum and a translation part t_{ij}, that is exhibited as a phase gradient. This is shown in Eq.(5). Suppose f_i and f_j satisfy

$$f_j(x + \triangle x, y + \triangle y) = f_i(x, y) \tag{4}$$

The Fourier transform of Eq.(4) yields the following by the *shift* theorem

$$F_j(\omega_x, \omega_y) \, e^{j(\omega_x \triangle x + \omega_y \triangle y)} = F_i(\omega_x, \omega_y) \tag{5}$$

This allows separate estimation of the translational offset and the linear transform, which dramatically reduces the search space. The overall process of Fourier based estimation is illustrated in Fig.2. The translational offset , t_{ij} is efficiently obtained using the well-known relation, deduced from Eq.(5), between the correlation and the Fourier transform as below [8].

$$\operatorname{argmax}_{x,y} \left\{ F^{-1} \left[\frac{F_i(\omega_x, \omega_y) \, F_j^*(\omega_x, \omega_y)}{|F_i(\omega_x, \omega_y)||F_j^*(\omega_x, \omega_y)|} \right] \right\} \tag{6}$$

where $\tilde{f}_i$ represents an affine transformed version of a source block, f_i, and $*$ is the complex conjugate. Provided that the blocks are *prewhitened* (the mean is removed), a contrast function can be introduced as follows.

$$\hat{f}_j(\mathbf{u}) = \alpha_{ij} f_i(T_{ij}(\mathbf{u})) + \beta_{ij} \tag{7}$$
$$\alpha_{ij}^2 = \sum f_i^2(\mathbf{u}) / \sum f_j^2(\mathbf{u}) \quad \beta_{ij} = E[f_j(\mathbf{u})]$$

where α_{ij} is a contrast coefficient and β_{ij} is the mean of the target block. Various methods of reducing the amount of computation involved in estimating the linear transform (shaded part of Fig. 2) have been suggested. Three approaches that estimate the linear part of the geometric transformation are reviewed, and their strengths and shortcomings are identified. Particular attention is given to the work of Hsu, Wilson and Bhalerao [4], [6].

A. *Fourier Mellin Estimation*

This algorithm is designed for efficient estimation of scale and rotation based on the phase-correlation relation. An image rotation shifts the function $|F_j(\theta, r)|$ along the angular axis. A scaling of the image is reduced to a scaling of the radial coordinate and a magnification of the intensity by a constant factor r^2. Scaling can be further reduced to a translation by using a logarithmic scale for the radial coordinate.

$$\begin{aligned} F_i(\theta, \lambda) &= f_i(\theta, \log r) \\ F_j(\theta, \lambda) &= s^{-2} f_i(\theta - \theta_0, \log r - \log s) \end{aligned} \tag{8}$$

In this polar-logarithmic representation, both rotation and scaling are reduced to translation.

$$F_j(v, w) = s^{-2} e^{-j2\pi(v \log s + w\theta_0)} F_i(v, w) \tag{9}$$

Rotation and scale can be efficiently estimated by the coordinate of the maximal value of Eq.(6). This technique decouples image rotation, scaling, and translation, and is accurate and computationally efficient. However, it excludes the estimation of shear [9], [5].
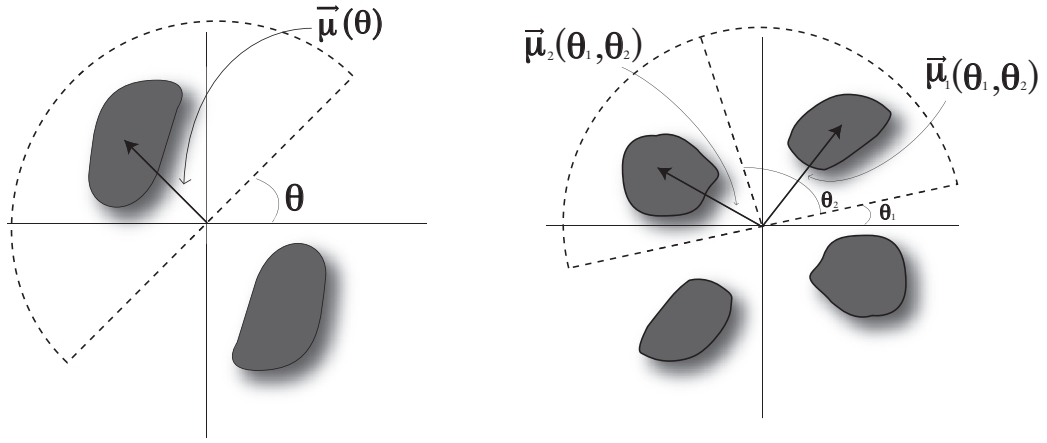
Fig. 3. Single cluster model with centroid $\vec{u}(\theta)$ (left) and double cluster model with centroid pair $\vec{u}_i(\theta_1, \theta_2)$

B. Angular Variance Analysis

The affine transform in Eq.(1) can be described as the alignment of a pair of 2D vectors to another pair. Assuming the presence of double energy clusters in the Fourier spectra of respective blocks, as shown in Fig.3, identifying representative centroids (affine invariant coordinates) of each cluster can disclose an affine transformation relating the two blocks by simply solving a matrix equation. Hsu and Wilson [4] introduced Angular Variance Analysis (AVA) to locate two representative centroids. This is conducted to find the angles, θ_1 and θ_2, which minimise the sum of the variances, $\sigma^2(\theta_1, \theta_2)$.

$$\sigma^2(\theta_1, \theta_2) = \frac{1}{E_{half}} \sum_{x \in \Lambda_{(\theta_1, \theta_2)}} |x| \cdot ||x - \mu||^2 \tag{10}$$

where $\Lambda_{(\theta_1, \theta_2)}$ denotes the coordinate set of the half plane subdivision defined by angles θ_1 and θ_2. μ is the representative centroid of the corresponding subdivision. E_{half} is the sum of the energy (magnitude) of the half plane. The angular segments Λ and the corresponding centroids μ are *equivariant* under the invertible transformation of Eq.(4), which makes the variance analysis legitimate. This process involves computation of the centroid and variance for every pair of angular segments to find θ_1 and θ_2 that minimise the variance sum. It requires considerable computation, but a fast algorithm using the partial summation technique was developed by Kruger and Calway [10]. A linear part of the affine matrix is estimated by simply aligning the centroids of the segments from both blocks, *i.e.* solving the following equation for a matrix, A_{ij}.

$$A_{ij} \cdot M_i = M_j \tag{11}$$

where M_i is a 2×2 matrix that consists of a centroid pair, (μ_1^T, μ_2^T) obtained from block, f_i. Due to the Hermitian symmetry of the spectra, it is necessary to examine all eight possible centroid alignment combinations and choose the one with the maximum correlation using Eq.(6).

C. Gaussian Mixture / Levenberg Marquart Optimisation

Bhalerao and Wilson [11], [6] approached the estimation of the affine transformation using least square optimisation, the Levenberg-Marquardt (LM) optimisation with a Gaussian mixture model.

The Levenberg Marquart (LM) method [12] is a non-linear data fitting algorithm and has become the standard of nonlinear least square routines. One can imagine this method as an extension of the simple Newton method. An important feature of this method is that the Hessian matrix can indicate how far the slope extends, whereas the Jacobian matrix can tell only the gradient. A factor, λ is introduced to control the diagonal dominance of the Hessian matrix of second order partial derivatives of unknowns. Increasing

or decreasing the factor allows the position to move back and forth on the surface of least-squares. In other words, if the least-square error increases, it is possible to change direction by the increasing the factor, decreasing it otherwise. This guarantees convergence unlike the Newton algorithm.

Bhalerao and Wilson's algorithm proceeds as follows.
1) Estimate the Fourier power spectrum with a Gaussian mixture model using LM
2) Estimate the affine parameters using LM

A Gaussian Mixture Model (GM) [13] is a type of probability density model which combines an arbitrary number of Gaussian models. Each constituent Gaussian model is weighted by a factor ω_i. Employing a Gaussian mixture model to represent the directional and symmetric distribution of the Fourier spectrum is an admirable approach in that each constituent Gaussian model G_i can effectively represent a distinctive directional energy cluster. To model the Fourier spectrum, a set of partial derivatives of the covariance matrix of the Gaussian mixture model is needed.

$$\frac{\partial G_i(u)}{\partial \omega_m} = G_i(u)/\omega_m$$

$$\frac{\partial G_i(u)}{\partial \Sigma_m^{-1}} = G_i(u)uu^T/2 \tag{12}$$

The LM then fine-tunes the covariance matrix of each Gaussian model such that the error between the mixture model and the power spectrum is minimised.

$$\Sigma_u(G_i(u) - |F_i(u)|)^2 \tag{13}$$

In the same manner, the LM algorithm optimises the affine parameters so that the overall error shown in Eq.(14) is minimised using partial derivatives of the parameters.

$$\Sigma_u(A_{ij}(G_i(u)) - |F_j(u)|)^2 \tag{14}$$

$$\frac{\partial G_i(A(u))}{\partial A} = -G_i(A(u)) \sum_m^M \Sigma_m^{-1}Auu^T \tag{15}$$

D. Remarks

The Fourier-Mellin based method is an efficient and accurate estimation technique where the advantages are maintained even in the presence of noise. The method has been deployed in many applications such as image matching, recognition, and registration, and continues to gain widespread support in new areas. The estimation, however, excludes *shear* deformation, which is essential in a viewpoint related task that involves skewed images.

The advantage of Angular Variance Analysis is that the full affine transform is computed by simply aligning the centroids. The segments defined in terms of variance help the centroids latch onto the significant image structure. Nevertheless, the approach makes the implicit assumption that double distinct energy clusters divide the half-plane. This may produce a less accurate result when more that two clusters are present or when the distinctiveness of the energy clusters is less obvious. The reason is that the centroid fails to latch onto the same structure in respective spectra. The approach may easily be extended to deal with an arbitrary number of clusters by dividing the half-plane accordingly, but it still lacks the ability to decide on the number of clusters. Calway [7] introduced a simple metric to determine the underlying cluster model for the spectrum: $\sigma_{\theta_1}^2 + \sigma_{\theta_2}^2/\sigma_{(\theta_1,\theta_2)}^2 < t_\sigma$ indicates which model in Fig.3 better represents the spectrum. However it only works well with up to double clusters and more work is required to determine the threshold, t_σ in an adaptive manner.

The least-square optimisation approach is simple and powerful, and easily overcomes the problems identified in the previous method [4]. However, the computational requirement is significant. This is mainly because the error shown in Eq.(14) needs to be updated at every iteration in order to guide the direction of the gradient of the least-squares. Secondly, the computation increases dramatically as more Gaussian components are used. Modelling with an insufficient number of components hinders accurate approximation, and therefore the accuracy of the estimation decreases. Thirdly, by the nature of the nonlinear optimisation, the method cannot escape from local optima, as illustrated. The resulting parameters vary depending on the initial guess, some of which lead to local optima. Having multiple initial guesses attenuates the problem but the computational burden increases still further. A new approach that partly combines the previous three methods is presented in the next section.

III. A Combined Gaussian Mixture / Angular Analysis Approach

A combined solution is suggested that acknowledges the efficiency of direct affine estimation and the adaptiveness of the Gaussian mixture modelling strategy. The algorithm proceeds as follows:

1) The spectral data is modelled by a Gaussian mixture using Expectation Maximisation (EM).
2) The resulting model undergoes an angular analysis that discloses the direction and strength of an arbitrary number of features
3) One (or two) components are selected as the representative Gaussian models to be aligned (corresponding to the centroid in the AVA).

In [3], the EM was modified so that the analysis is also performed during the estimation iteration but it was realised that this hinders fast convergence. The chosen representative Gaussians are used in Eq.(11) to estimate the affine transformation.

A. Gaussian Mixture Modelling

Expectation Maximisation (EM)[13] is an estimation algorithm driven by the probability of each data point belonging to a class. EM is a favourite choice for statistical analysis due to its simplicity and fast convergence rate. Estimation is performed by repeating two steps, the E-step and the M-step. An example with respect to the Gaussian mixture model follows. The E-step, computes expected classes for all data points for a given class and is given by

$$P(\omega_i|x_k,\mu_t,\Sigma_t) = \frac{p(x_k|\omega_i,\mu_i(t),\Sigma_i(t))\omega_i(t)}{\sum_{j=1}^{c} p(x_k|\omega_j,\mu_j(t),\Sigma_j(t))\omega_j(t)} \tag{16}$$

where Σ_i, μ_i, w_i, and t denotes a covariance matrix, centroid, weight of Gaussian G_i and iteration respectively. In the M-step, the maximum likelihood is estimated given the new membership distribution of the data

$$\begin{aligned}
\mu_i(t+1) &= \frac{\sum_k P(\omega_i|x_k,\mu_t,\Sigma_t)x_k}{\sum_k P(\omega_i|x_k,\mu_t,\Sigma_t)} \\
\Sigma_i(t+1) &= \frac{\sum_k P(\omega_i|x_k,\mu_t,\Sigma_t)(x_k-\mu_i(t+1))(x_k-\mu_i(t+1))^T}{\sum_k P(\omega_i|x_k,\mu_t,\Sigma_t)} \\
\omega_i(t+1) &= \frac{\sum_k P(\omega_i|x_k,\mu_t,\Sigma_t)}{\text{number of data points}}
\end{aligned} \tag{17}$$

Having an insufficient number of Gaussian components and/or incorrect initial guesses always results in a local optima problem, to which EM is susceptible. To alleviate this, at least six Gaussian components of the initial guess are distributed (0 to π) with a uniform angular interval.
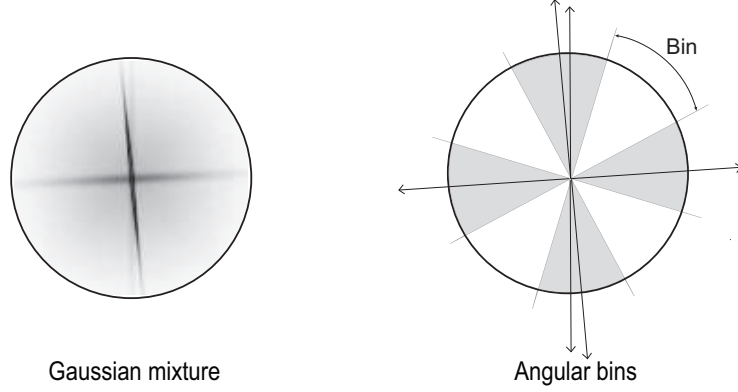
Fig. 4. Gaussian mixture model and angular bins

B. Selection of Representative Gaussian Model

In order to determine two significant Gaussian components, a polar bin grid is adopted for angular analysis as shown in Fig.4. Each bin is defined and populated as follows.

$$B_i = \{G_j | \frac{(2i-1)\pi}{2k} + \theta(G_{max}) < \theta(G_j) \le \frac{(2i+1)\pi}{2k} + \theta(G_{max})\} \tag{18}$$

where $(0 \le j \le l), (0 \le i \le k-1)$ and $\theta(G_j)$ is the direction of the principal eigenvector of a Gaussian component, G_j, and G_{max} is a Gaussian component with the highest weight, ω_{max}. l is the number of Gaussians in the mixture. k is set to $imagesize \times 3/8$, which is found empirically. In each bin, a Gaussian with the greatest weight is selected. Some bins may be empty and some bins may contain only Gaussians with a negligible weight, which are treated as empty. Counting the number of selected Gaussian components (or non-empty bins) shows the number of strong directional features. If the number of features is greater than that of the target block or if more than two features are found, the one with the smallest weight is discarded, and this is repeated until the numbers match. Once the required number of strong Gaussian components is determined, a coordinate, $\hat{\mu}$ of the representative slices is calculated for alignment as follows, where p is fixed for all (*e.g.* $p = 0.5$).

$$\hat{\mu}_i = \sqrt{-2\sigma_i^2 \log_e p} \ \cdot \vec{G}_i \tag{19}$$

where $0 < p \le 1$ and σ_i^2 and $\vec{G}$ are the variance (eigenvalue) and the eigenvector of a principal component of a Gaussian, G_i respectively. The found coordinates, $\hat{\mu}$ are used for alignment using Eq.(11).

C. Discussion

The suggested approach benefits from the efficiency of the direct affine estimation and the adaptive modelling of the Gaussian mixture, in particular it can reveal the number of significant directional features and determine the two stronger features if more than two are found. In the case where the number of features in the source and target patches does not match, existing methods result in a poor estimation. The ability to determine the number of features allows us to force the estimation to be based on the lower number of features. This can prevent an extreme shear transformation being wrongly estimated, see Fig.5. The estimation process is still expensive and time consuming compared with AVA, because of the iterative nature of the EM. In the next section, a deterministic approach that reduces the estimation time while retaining the important attributes is presented.
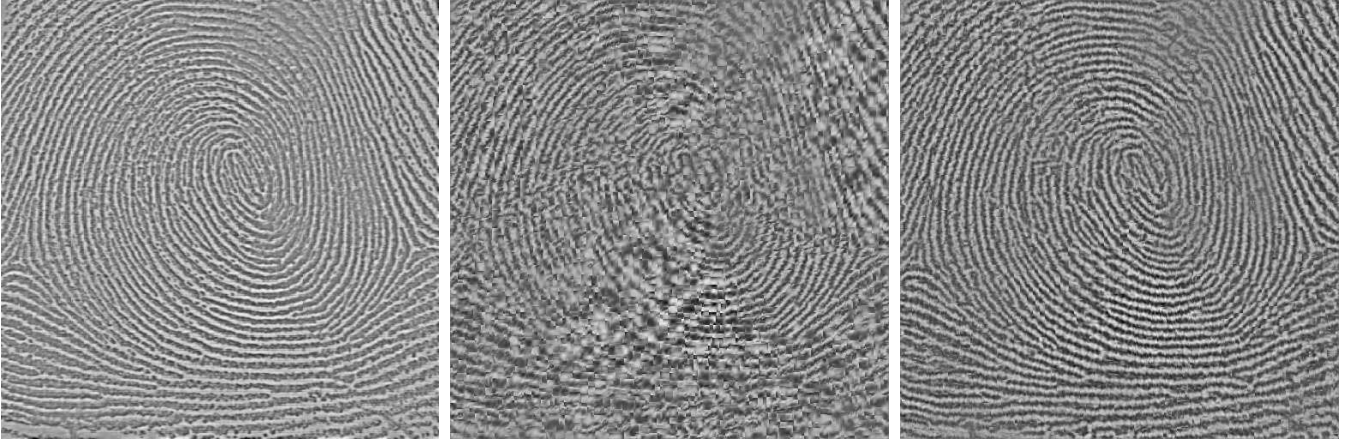
Fig. 5. Approximation experiment using `fingerprint`: (from left to right))original image, reconstruction by two-component method with Calway's single / double cluster model, reconstruction by the modified EM for GMM

IV. FOURIER SLICE BASED AFFINE ESTIMATION

AVA has merits in direct estimation but it may work less effectively when the underlying cluster model does not suit the spectrum. Optimisation based estimation approaches the problem by minimising the error of a least-square parametric equation iteratively between the spectrum and the mixture model (the more components, the more accurate the modelling). The latter is a more robust solution than the cluster model, however the computational requirement is prohibitive. An experimental algorithm is presented that, albeit slow, combines the beneficial parts of the previous methods. A more efficient and accurate algorithm based on Fourier slice analysis is then proposed.

An accelerated variant of the previous algorithm is suggested that uses a similar strategy to the Gaussian mixture based angular analysis. The computational requirement is greatly reduced by substituting a deterministic step for the EM. The iterative process of the EM is not desirable in some real-time applications or for hardware implementation. An analysis, analogous to a Gaussian mixture, is performed by considering an angular slice through the origin of the Fourier spectrum, the *Fourier slice*. Using the Fourier slice also allows us to utilise the slice-projection theorem, by which the phase gradient can be taken into account, hence resulting in better estimation. The algorithm proceeds as follows.

1) Extract the polar contour signature C using Eq.(20)

$$
\begin{aligned}
C &= \{(r, \theta) \,|\, r = S(\theta), 0 \leq \theta \leq \pi\} \\
S(\theta) &= \sum_x \sum_y |F_C(x, y)| \delta(x \cos \theta + y \sin \theta)
\end{aligned}
\tag{20}
$$

2) Perform angular binning analysis to select a representative Fourier slice.
3) Perform spectral alignment with a refinement step of slice-wise phase correlation.

The polar contour signature of the energy distribution is illustrated in Fig.6.

A. Selection of Representative Slice

Angular analysis, similar to Fig.4 is performed to find the significant slices as illustrated in Fig.6, with $\omega(\theta)$ that corresponds to the weight, ω of the mixture model.

$$
\omega(\theta) = \frac{r(\theta)}{\sum_{t=0}^{\pi} r(t)}
\tag{21}
$$

The remaining process is identical to that of the previous methods using Eq.(11) except that the centroid, μ_θ of the slice is determined with only half the slice, one end of which is the origin, as below.

$$
\begin{aligned}
\mu_\theta &= (E(H_\theta) \cos \theta, E(H_\theta) \sin \theta) \\
H_\theta &= \{S(r, \theta) | 0 \leq r \leq M/2\}
\end{aligned}
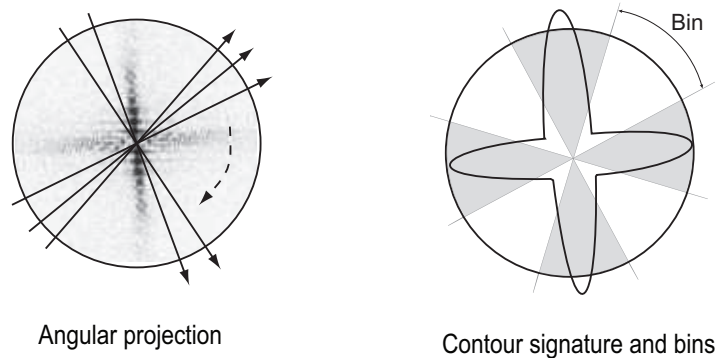\tag{22}
$$

Angular projection

Contour signature and bins

Fig. 6. Fourier slice and angular bins

where M denotes the size of Fourier slice.

B. Slice-wise Phase-correlation

As a refinement of estimation at the expense of a slight increase in computation, the phase correlation can be taken into account in the estimation of the linear component. Eq.(21) resembles that of the Radon transform. The Fourier slice that passes through the origin indeed has a mathematical relation to the Radon transform as below, the *projection slice theorem*.

$$Rf(\theta) = \tilde{F}(\theta) \tag{23}$$

where $\tilde{F}$ is the Fourier slice function and R is the Radon projection in the spatial domain. The theorem states that a Fourier slice at angle θ, is the Fourier transform of the Radon projection at angle θ. As slices of the Fourier spectrum of images are readily available, it is possible to simply apply phase correlation using Eq.(6) to find a slice with the highest *goodness of fit* (maximal correlation) in the target image and rectify the representative slice. That is, to compute the correlation of the slices in the vicinity ($\pm\theta_\epsilon$) of the representative slice with the representative slice of the target block f_j, and then to choose a new slice with the greatest correlation as the representative slice. Care should be taken when applying the phase correlation. Either both slices are normalised using Eq.(8) so that the centroids of the slices latch onto the same coordinate, or the correlation should be based on a logarithmic scale to cope with scale difference using Eq.(9), where search for the maximum correlation is confined to a range corresponding from 0.5 to 2 in terms of the scaling factor. This process improves the accuracy of the estimation (in particular for *shears*) by comparing the spatial structure.

C. Implementation

There are a couple of issues to be addressed regarding implementation. The slice-wise phase correlation operation requires several direct and inverse Fourier transforms. The discrete implementation must be done with care to avoid artefacts due to sampling and truncation. In the case of the unavailability of the polar Fourier spectrum, the Radon projection of an image in the same orientation followed by the Fourier transform provides the required Fourier slice, by Eq.(23). It is helpful to window the images before calculating the Fourier transforms. This effectively removes the *cross sign* from the power spectrum that is caused by the boundary wrapping of the source. The $\cos^2$ window function $w(y)$ was used, as given by Eq.(24).

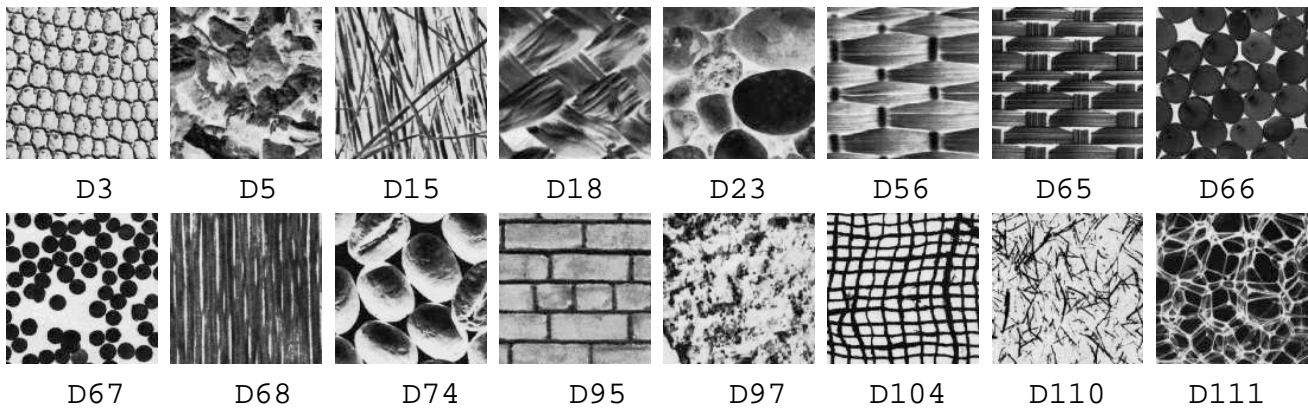$$w(y) = cos^2[\pi p/2B]cos^2[\pi q/2B] \tag{24}$$
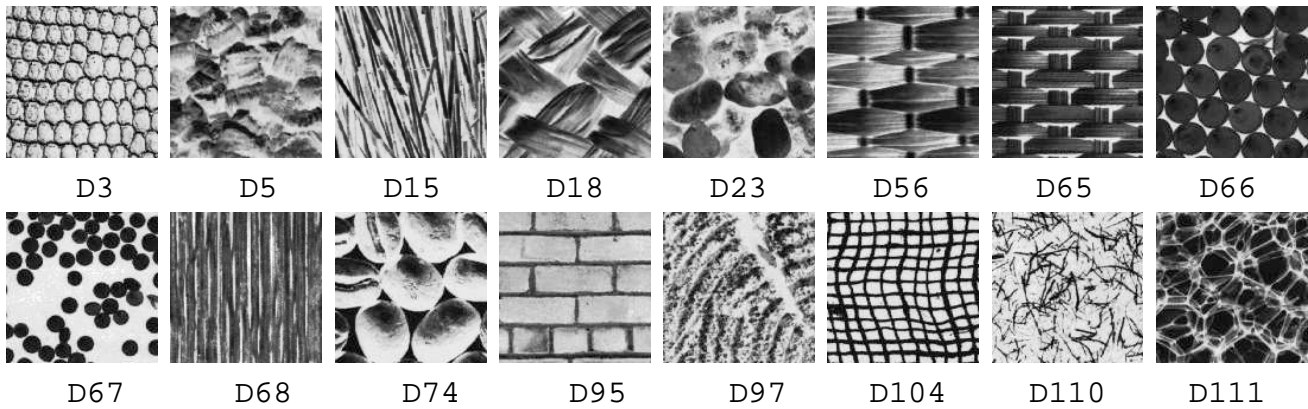
Fig. 7. *Brodatz* texture samples

Fig. 8. *Brodatz* texture samples for approximation test

Two issues arise in the logarithm resampling of the spectral magnitude of the image to polar-log coordinates. Care needs to be taken in selecting the starting point of the logarithmic resampling, due to $\log 0 = \infty$. The rotation angle and scaling factor can be estimated with an accuracy that depends on the sampling accuracy, where sub-pixel precision generally results in a more accurate location of the correlation maximum. The angular coordinate of the spectral magnitude is sampled uniformly and the precision of the rotation angle is therefore uniform. However, the logarithmic distortion along the radial coordinate results in a non-uniform precision in scale estimation. The scaling estimation error becomes large as it increases or decreases away from $r = 1$. Experimental evaluation showed that no significant error is incurred in the range of $0.5 \le r \le 2$ with a half-pixel precision.

V. EXPERIMENTAL EVALUATION

To illustrate the effectiveness of the algorithms, results are presented for transformation estimation involving different texture types as shown in Fig.7. The texture images are of size 128×128 pixels and are corrupted by additive, uncorrelated Gaussian noise to give a SNR of 10dB. The images are pre-filtered using a highpass filter with bandwidth proportional to the block size, in order to deal with pure texture using the MFT [14]. The number of angular segments in the AVA is set to the $imagesize$ of 128. The texture images are bandpass filtered by taking the two lowest levels (high frequency components) from the Laplacian pyramid. The reason to remove low frequency components is to focus on the pure texture information which generally resides in the high frequency band, and to remove the distracting energy in

TABLE I
COMPARATIVE PSNR RESULTS OF *rotation+scale* TRANSFORMATION (IN DB)

sample	F-Mellin	V-Analysis	L-Optmiz	M-Angular	F-Slice
D3	24.53	24.35	24.52	24.43	**24.57**
D5	26.89	25.43	26.49	**27.37**	26.11
D15	**27.33**	25.35	25.64	**27.33**	26.50
D18	27.39	25.34	25.51	26.24	**27.52**
D23	26.12	23.34	25.24	**27.45**	25.30
D56	26.19	26.43	25.64	**27.37**	26.44
D65	26.03	25.75	25.75	26.17	**26.42**
D66	**27.44**	25.21	25.77	27.34	26.74
D67	**27.59**	25.04	25.26	27.56	26.64
D68	**26.92**	25.43	25.73	26.73	26.32
D74	**27.44**	24.92	25.71	27.23	26.67
D95	27.17	25.43	25.67	**27.75**	26.57
D97	26.51	24.88	25.73	26.21	**26.73**
D104	26.02	25.54	25.45	26.34	**26.46**
D110	26.49	24.70	25.13	**26.75**	26.23
D111	26.18	25.21	25.73	**27.32**	26.38
time(sec)	2	4	21	20	2

TABLE II
COMPARATIVE PSNR RESULTS OF *rotation+scale+shear* TRANSFORMATION (IN DB)

sample	F-Mellin	V-Analysis	L-Optmiz	M-Angular	F-Slice
D3	22.27	24.80	24.52	24.15	**25.12**
D5	25.36	26.25	26.49	27.16	**28.53**
D15	23.29	26.22	26.65	**27.34**	26.75
D18	24.76	27.60	**27.73**	27.21	27.23
D23	22.10	24.12	25.22	**27.45**	23.50
D56	27.83	**28.74**	28.66	27.16	28.64
D65	27.72	27.41	27.84	27.17	**28.01**
D66	25.57	26.75	**25.75**	27.04	26.00
D67	25.26	25.13	25.75	26.56	**27.34**
D68	23.26	27.66	27.81	27.57	**27.89**
D74	22.21	25.98	**27.84**	27.31	26.88
D95	23.04	25.32	26.43	27.42	**28.00**
D97	24.12	25.53	25.67	27.21	**28.04**
D104	22.80	27.53	27.43	27.57	**28.06**
D110	23.30	27.27	27.75	27.89	**28.90**
D111	25.73	27.85	27.33	27.09	**28.23**
time	2	4	21	20	2

the Fourier spectrum for accurate estimation of the energy clusters.

The first investigation tests the accuracy of the estimated transform as well as the noise robustness. The fidelity of the warping approximation is measured by the PSNR of the approximated texture, $\hat{f}_j$, which is warped by the estimated transform. The intensity scaling estimation is ignored ($\alpha = 1$) to avoid any bias that may affect the affine estimation results. Table.I shows the comparative experimental results of five algorithms; Fourier-Mellin (F-Mellin), Angular variance analysis (V-Analysis), Levenberg-Marquart (L-Opimiz), Gaussian mixture / angular analysis (M-Angular), and Fourier slice (F-Slice). The true affine transform is rotated by $\pi/3$, followed by a scaling of $x : 1.2$ and $y : 1.0$. It is noticeable that the AVA provides a less effective estimation for multi-directional textures such as D5, D23, D67, D74, D97 and D110. This confirms that using an adaptive model to better represent the data is an important factor.

The second evaluation tests the previous deformations when combined with a shearing of $x : -0.3$ and $y : 0.9$. The results are shown in Table.II. It is observed that Fourier slice analysis provides better warping

TABLE III
COMPARATIVE PSNR RESULTS OF TEXTURE *approximation* (IN DB)

sample	F-Mellin	V-Analysis	L-Optmiz	M-Angular	F-Slice
D3	20.24	19.51	22.15	23.11	**24.34**
D5	14.67	15.49	16.16	18.53	**19.75**
D15	19.11	20.65	20.34	**21.75**	21.11
D18	19.83	21.73	21.11	21.23	**22.75**
D23	19.47	19.22	19.45	**19.50**	19.12
D56	**20.23**	20.21	20.06	20.14	20.20
D65	26.62	26.84	26.17	**26.71**	26.41
D66	19.07	20.55	17.04	19.00	**21.75**
D67	17.93	17.75	16.56	16.34	**18.13**
D68	**27.73**	27.55	27.73	27.57	**27.73**
D74	17.43	19.84	19.31	19.88	**19.98**
D95	**25.48**	25.35	25.37	25.33	25.31
D97	13.98	15.67	17.11	18.04	**20.53**
D104	18.24	18.43	18.53	18.06	**20.57**
D110	20.91	19.17	18.75	20.89	**21.90**
D111	21.12	19.85	**22.33**	20.09	22.23
time	2	4	21	20	2

results for transformations involving *shear*. This is attributed to the phase correlation in the linear part estimation.

The final experiment tests the approximation of a texture patch to a similar but non-identical texture patch, sampled from the same texture image but from different parts as shown in Fig.8. Intensity scaling α is enabled in this evaluation. Table.III shows the results. The Fourier slice based estimation provides generally better estimation with good computational efficiency.

Being self-similar, natural images contain considerable amounts of redundancy. Modelling by affine symmetry the self-similarity that exists between subblocks offers a different viewpoint from which to address various tasks [11]. This model resembles fractal theory as only a single prototypical block is needed to replicate other blocks of similar texture [3]. In the same manner, the Warplet [4], [6] could be applied to classify sub-blocks of an image so as to perform segmentation. Experimental image approximation is illustrated in Fig.9.

VI. CONCLUSIONS

A general purpose block-to-block affine transform estimation method has been presented. It has use in many applications, including image segmentation, registration, image coding and motion analysis. Motion estimation [2], [15] and image registration [16], [17], [18], [19], [1], in particular, require fast computation and high reliability. Experimental results show that, compared with existing methods, the proposed technique not only reduces the computational burden but also increases the reliability of the estimation and maintains noise robustness.

The proposed technique can be summarised as follows.
1) Identify the significant directional structure from the power spectrum, which is invariant under geometric deformation by the *affine theorem*
2) Apply slice-wise phase correlation to determine the exact corresponding invariant points.
3) Finally perform spectral alignment to reveal the affine transformation.

The key elements of these improvements are three-fold. The first is the use of angular bins that determine the number of significant directional features and change the estimation strategy accordingly. The second is the deterministic nature of the Fourier slice-projection calculation that replaces the iterative process
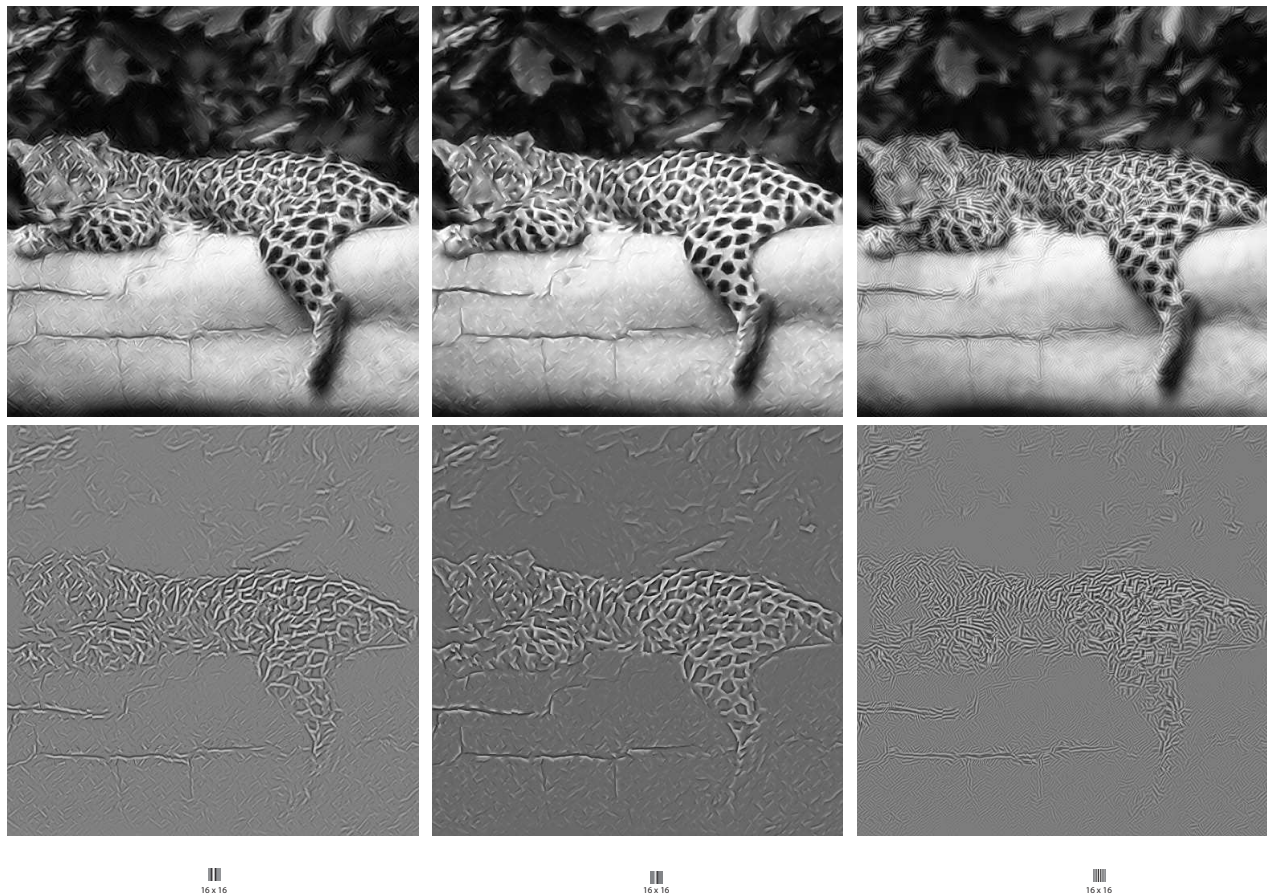
Fig. 9.   Use of various prototypes: `jaguar` *(continued)*; (top)line (middle) cliff, (bottom) ridge (left) highpass approximation (right) image reconstruction

of mixture modelling. The last is taking phase information into account in the spectral alignment. This improves the representative points on which identical structure in the two images can be latched.

The proposed technique shows very encouraging performance in terms of estimation accuracy, and the computational requirement remains comparable to the accelerated variant of angular variance analysis. The improvements are important to applications that require estimation of the affine transform, for example in fast motion estimation. Optimisation is needed in many parts of the implementation to further reduce the computation time. For example, conversion of the Fourier transform to a polar representation takes a significant percentage of the overall processing time, when matching one slice to another in the target block. Computing the correlation function for a subset of candidate slices judged by the energy distribution around the current representative slice rather than computing it for all possibilities is another means of reducing complexity.

REFERENCES

[1] S. Krüüger and A. Calway, "Image registration using multiresolution frequency domain correlation," in *BMVC.*, 1998, pp. 316–325.

[2] R. Wilson, P. Meulemans, A. Calway, and S. kruüger, "Image sequence analysis and segmentation using G-blobs," in *IEEE ICIP.*, 1998, pp. 483–487.

[3] 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.

[4] 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.

[5] Y. Keller, A. Averbuch, and M. Israeli, "Pseudopolar-based estimation of large translations, rotations, and scalings in images," *IEEE Trans. Image Process.*, vol. 14, no. 1, pp. 12–22, Jan. 2005.

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

[7] A. D. Calway, "Image representation based on the affine symmetry group," in *IEEE ICIP.*, Sep. 1996, pp. 189–192.

[8] I. Pitas, *Digital Image Processing Algorithms*. Prentice Hall, 1993.

[9] Q. Chen, M. Defrise, and F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 16, pp. 1156–1168, 1994.

[10] S. Kruüger and A. Calway, "Multiresolution motion estimation using an affine model," University of Bristol, Techical Report CSTR-96-002, June 1997.

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

[12] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 2000.

[13] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[14] 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.

[15] H. Park, A. C. Yu, and G. Martin, "Progressive mesh-based motion estimation using partial refinement," *LNCS*, vol. 3893, pp. 98–101, Apr. 2006.

[16] D. Barnea and H. Silverman, "A class of algorithms for fast image registration," *IEEE Trans. Comput.*, vol. 21, no. 2, pp. 179–186, 1972.

[17] G. Vanderburg and A. Rosenfeld, "Two-stage template matching," *IEEE Trans. Comput.*, vol. 26, pp. 384–393, 1977.

[18] A. Goshtasby, "Template matching in rotated images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 7, no. 3, pp. 338–344, 1985.

[19] J. Segman, "Fourier cross correlation and invariant transformations for an optimal recognition of functions deformed by affine groups," *J. Opt. Soc. Am. A.*, vol. 9, no. 6, pp. 895–902, 1992.