Supervised Motion Segmentation by Spatial-Frequential Analysis and Dynamic Sliced Inverse Regression

Han-Ming Wu and Henry Horng-Shing Lu

National Chiao-Tung University, Taiwan

Abstract: In this paper, we propose a new method for supervised motion segmentation based on spatial-frequential analysis and dimension reduction techniques. A sequence of images could contain non-ridge motion in the region of interest and the segmentation of these moving objects with deformation is challenging. The aim is to extract feature vectors that capture the spatial-frequential information in the training set and then to monitor the variations of the feature vectors over time in the test set. Given successive images in the training set, we consider a dynamical model that extends the sliced inverse regression in Li (1991). It is designed to capture the intrinsic dimension of feature vectors that holds over a local time scale. These projected features are then used to classify training images and predict forthcoming images in the test set into distinct categories. Theoretic properties are addressed. Simulation studies and clinical studies of a sequence of magnetic resonance images are reported, which confirm the practical feasibility of this new approach.

Key words and phrases: Motion segmentation, non-ridge motion, spatial-frequential analysis, dimension reduction, Gabor filter bank, sliced inverse regression.
Running title: Motion Segmentation by DSIR

1 Introduction

Motion segmentation is a preliminary step for further analysis of dynamic images, like videos, medical images, satellite images, and so forth. It is aimed to partition the images into regions that have different motion characteristics. Non-rigid motion is the motion of objects that change shape (i.e., deform) over time. Segmentation from these temporal sequences of images became difficult and challenging.

There are various approaches proposed for motion segmentation in literature by statistical modelling (e.g., Bouthemy, 1989, Nguyen, Worring, and Dev, 2000, Vasconcelos, and Lippman, 2001) and/or structural modelling (e.g., Duncan, Owen, Staib, and Anandan, 1991, Isard and Blake, 1996, Mikic, Krucinski, and Thomas, 1998, Deng and Manjunath, 2001). Most of the above methods need to specify and estimate the parameters for their models of motion, which are either difficult to determine or computational demanding in estimation procedures. The computational overhead could be overcome by the increasing power of computers. However, as the computer power increases, the volume of data and images collected also increases. We still need to find fast and effective methods to mine important features and intrinsic dimension.

Inspired by the recent studies of the early vision system of human perception in V1 cells, image synthesis and segmentation of static images based on spatial-frequential analysis that mimics human vision were investigated in literature (Malik and Perona, 1990, Dunn, Higgins, and Wakeley, 1994, Tan, 1995, Zhu, Wu and Mumford, 1998). Recent progresses have been made for the modelling of textures and dynamic textures by statistical models (Zhu, Liu, and Wu, 2000, Wu, Zhu, and Liu, 2000, Soatto, Doretto, and Wu, 2001, Wu, Zhu, and Guo, 2002). It remains challenging to segment dynamic images with only a few training images. We are thus aimed to extend the studies of spatial-frequential analysis for motion segmentation.

Sliced inverse regression (SIR) is proposed by Li (1991) to find compact representations that explore the intrinsic structure of high-dimensional observations. It has been extended and used in various applications (Chen and Li, 1998, Li, 2000). We have applied this technique with spatial-frequential analysis to segment and diagnosis static images, like ultrasound images (Chen, Lu, and Lin, 1999, Chen, Lu, and Han, 2001, Lu, Chen, and Wu, 2001). This study will further investigate the possibility of extending SIR and spatial-frequential analysis for dynamic images.

We will first extend the model of SIR for dynamic data and the new model is called as dynamic SIR (DSIR). Then, DSIR is combined with spatial-frequential analysis for motion segmentation. Every pixel of an image is regarded as an realization of a stochastic process over space and time. The feature vector for one pixel in one time frame is first analyzed by spatial-frequential analysis of local blocks centered at that pixel. Assume that the relationship between these feature vectors and class labels remains similar between successive frames for neighboring pixels, then we can find out the intrinsic dimension of feature vectors in the training images by DSIR. These projected feature vectors thus provide prediction rules for forthcoming images in the test set. Furthermore, we will only need a small number of training images to decide the projection of feature vectors and prediction rules.

The model setup of DSIR and its properties are given in Section 2. Section 3 discusses the feature extraction methods. The algorithm of DSIR and the segmentation procedure are presented in Section 4. Section 5 reports the segmentation results both on the simulated and clinic image sequence of magnetic resonance (MR) images. Comparison studies are also performed to investigate the projection directions by the principal component analysis (PCA), independent component analysis (ICA), SIR and DSIR. We then conclude in Section 6. The proof of the theoretic result is in the Appendix 1. All figures and tables are put in Appendix 2.

2 Models and Properties

Li (1991) introduced the following model

$$y = f(\beta_1'\mathbf{x}, \cdots, \beta_K'\mathbf{x}, \epsilon), \tag{1}$$

where y is a univariate variable, $\mathbf{x}$ is a random vector with dimension $p\times$ 1, $p \geq K$, β's are vectors with dimension $p\times$ 1, ϵ is a random variable independent of $\mathbf{x}$, and f is an arbitrary function. The β's are referred to effective dimension reduction (*e.d.r.*) or projection directions. Sliced inverse regression (SIR) is a method for estimating the *e.d.r.* directions based on y and $\mathbf{x}$. Under regular conditions, it is shown that the centered inverse regression curve $E[\mathbf{x}|y]-E[\mathbf{x}]$ is contained in the linear subspace spanned by $\beta_k\Sigma_{\mathrm{XX}}$ ($k = 1, \cdots, K$), where Σ_{XX} denotes the covariance matrix of $\mathbf{x}$. Based on these facts, the estimated β's can be obtained by the procedures of standardizing $\mathbf{x}$, partitioning slices (or groups) according to the value of y, calculating the slice means of $\mathbf{x}$, and performing the principal component analysis of the slice means with weights.

The above model, (1), can be extended for dynamic data as follows:

$$y(t) = f(\beta_1'\mathbf{x}(t), \cdots, \beta_K'\mathbf{x}(t), \epsilon(t)), \tag{2}$$

where $y(t)$ and $\mathbf{x}(t)$ are response variables and p-dimensional covariates observed at time t. The projection directions, β's, are assumed to be invariant over time and $\epsilon(t)$ is the stochastic process of noise. Analogy to the steps in Li (1991, 2000), we can assume the following condition and prove the theorem. The proof is given in Appendix.

Condition 1 *For any b in R^p, $E[b'\mathbf{x}(t)|\beta_1'\mathbf{x}(t), \cdots, \beta_K'\mathbf{x}(t)]$ is linear in $\beta_1'\mathbf{x}(t), \cdots, \beta_K'\mathbf{x}(t)$ for any t. That is, $E[b'\mathbf{x}(t)|\beta_1'\mathbf{x}(t), \cdots,\ \beta_K'\mathbf{x}(t)] = c_0 + c_1\beta_1'\mathbf{x}(t) + \cdots + c_K\beta_K'\mathbf{x}(t)$ for some constants $c_0,\ c_1, \cdots,\ c_K$ and any t.*

Theorem 1 *Under Model (2) and Condition 1, $E[\mathbf{x}(t)|\ y(t)] - E[\mathbf{x}(t)]$ falls in the linear subspace spanned by $Cov(\mathbf{x}(t))\beta_k$ for any t, $k = 1, \cdots, K$.*

When $\mathbf{x}(t)$ is elliptically symmetric for any time t, the above condition is fulfilled. This condition is weaker than the assumption of elliptical symmetry of $\mathbf{x}(t)$. When this condition is violated, the biases in estimating the projection directions are not large as discussed in Li (1991, 2000). Because the model (2) does not put any structural modelling over time, we can pull together successive data to estimate the projection directions more effectively as long as the projection directions remain the same. Furthermore, all the existent properties for SIR can be passed to this dynamic model without any difficulty. We thus called this model as the dynamic SIR (DSIR) model. The algorithm for estimating projection directions in motion segmentation is elaborated in Section 4.

Li, Aragon, and Thomos-Agan (1995) and Li (2000) investigated the analysis of multivariate outcome data. They considered the following time series model for a univariate response:

$$y(t) = \mu(t) + c_1(\mathbf{x})f_1(t) + \cdots + c_L(\mathbf{x})f_L(t) + \epsilon(t), \tag{3}$$

where $y(t)$ is a response curve, $\mu(t)$ is a baseline curve, the coefficients $c_l(\mathbf{x})$, $l = 1, \cdots, L$, are time-invariant factors that depend only on the covariate $\mathbf{x}$, $\{f_1(t), \cdots, f_L(t)\}$ are basis function and $\epsilon(t)$ represents the stochastic process of noise. Our model (2) differs from the model (3) in three folds. Firstly, we allow the covariate process, $\mathbf{x}(t)$, changes along time but assume the projection directions holds for any time. Then, we do not assume the noise is additive. While our model allows nonlinearity in f, their model allows nonlinearity in $c_l(\mathbf{x})$ and $f_l(t)$. Therefore, our model is applicable for a sequence of data where the projection directions and the nonlinear relationship hold, which usually works for a short sequence. On the other hand, their model is useful to determine the coefficient functions of covariates and basis functions of time separately from a sequence of data. They have extended the model (3) to multivariate responses and covariates. Similarly, it is also possible extend the model (2) to multivariate cases.

3 Feature Extraction of Images

We will discuss the feature extraction of 2D images here. These methods can be extended to 3D or higher dimension easily.

Space domain: Local blocks. A rectangular lattice for a digital image in 2D of size $N \times M$ is denoted by

$$\mathcal{S} = \{(I, J) | 1 \leq I \leq N, 1 \leq J \leq M, I, J \in \mathcal{Z}\}.$$

The spatial characteristic of a pixel in a image is described by its neighboring pixels. So, a local block in one time frame with size $b \times b$ can be formed as a feature vector $\mathbf{x}^{(i)}(t_j)$ of the central pixel $i \in \mathcal{S}$, $i = 1, \cdots, n$ and $n = (M - b + 1)(N - b + 1)$, for each time point t_j, $j = 1, \cdots, m$. The dimension of the feature vector in the space domain is b^2. We will not include the pixels at different time frames into the feature vector because they may vary according to motion. For the same pixel i, the collection of feature vectors along the sequence of images, $\{\mathbf{x}^{(i)}(t_j), j = 1, \cdots, m\}$, represents the temporal variation of the features by motion along time.

Given a sequence of m training images, we have the class labels, $\{y^{(i)}(t_j), j = 1, \cdots, m\}$, and the feature vectors, $\{\mathbf{x}^{(i)}(t_j), j = 1, \cdots, m\}$. We need to find the projection directions of feature vectors for classification and prediction. If the number of training images, m, is bigger than the dimension of feature vectors, b^2, then it is feasible to estimate the projection direction in the dimension of b^2. However, for a short sequence of training images, m is not necessary bigger than b^2. Therefore, we need to borrow the information from neighborhood pixels.

Let $\mathcal{N}_q^{(i)}$ be the set of neighboring sites for pixel i with the q-order neighborhood system. For example, the first order neighborhood system is the 4-neighborhood system, where every interior site has four neighbors and the size of $\mathcal{N}_1^{(i)}$ is 5. There are eight neighbors for every interior site in the second order neighborhood system, which is also called as the 8-neighborhood system and the size of $\mathcal{N}_2^{(i)}$ is 9. Then, for pixel i, we can collect the neighboring feature vectors as the training set:

$$\mathcal{X}^{(i)} = \{\mathbf{x}_l^{(i)}(t_j), j = 1, \cdots, m, l \in \mathcal{N}_q^{(i)}\}, \tag{4}$$

where $i = 1, \cdots, n$ and $n = (M - b + 1)(N - b + 1)$. For instance, if $q = 2$, then the size of training set becomes $9m$, which is bigger than b^2 when $m > b^2/9$. This training set will be used to estimate the projection directions of feature vectors for pixel i by DSIR in Section 4.

Frequency domain: Fourier transform of local blocks. If the feature of an image is periodical over space, then the feature of a local block in the space domain can be transformed to the frequency domain by the Fourier transform. This transform will highlight the periodical pattern (Weaver, 1983). This can be performed by Fast Fourier Transform if the block size is of the power of 2. For a block with size $b \times b$, the two dimensional discrete Fourier transform can be expressed as

$$F(u, v) = \frac{1}{b^2} \sum_{x=0}^{b-1} \sum_{y=0}^{b-1} f(x, y) \exp\left[-\mathbf{i}2\pi(\frac{ux}{b} + \frac{vy}{b})\right],$$

where $\mathbf{i} = \sqrt{-1}, u, v = 0, \cdots, b - 1$. Because the image intensity is real valued, the Fourier transform is symmetrical about the center. By this symmetry, almost a half of FFT calculation is redundant. Therefore, the feature in the frequency domain consists of $|F(u, v)|$ with dimension $b^2/2 + 2$ if b is power of 2.

Space-frequency domain: Gabor filter banks of local blocks. Human vision has demonstrated its superior capacity in detecting boundaries of desired objects. We will use the vision model based on our previous works (Chen, Lu, and Lin, 1999, Chen, Lu, and Han, 2001). Other similar approaches could be applied as well. We will construct a neuroimage by convolving the observed image with a bank of specific frequency and orientation bands, such as a bank of Gabor functions. The general form of a Gabor function is given by

$$g(x, y) = \exp\{-[(x - x_0)^2 a_1^2 + (y - y_0)^2 a_2^2]\pi\} \exp\{-2\pi i[u_0(x - x_0) + v_0(y - y_0)]\},$$

and its Fourier transform is

$$G(u, v) = \exp\{-\frac{1}{\pi}[\frac{(u - u_0)^2}{a^2} + \frac{(v - v_0)^2}{b^2}]\} \exp\{-2\pi i[x_0(u - u_0) + y_0(v - v_0)]\}.$$

Each local block is convolved with a bank of Gabor filters with different orientations and frequencies. We employed the so-called the *G-vector* as the feature vector at pixel (I, J), which is computed by

$$G_V(I, J) = \{g_{pk}(I, J), g_{nk}(I, J); k = 1, \cdots, r\},$$

where $g_{pk}(I, J)$ and $g_{nk}(I, J)$ are the summations of the positive and negative values for the neuroimage that is the convoluted image with the kth Gabor filter. Thus, the dimension of feature vector is $2r$ in the space-frequency domain. For instance, we can consider a bank of $r = 3 \times 8 = 24$ Gabor filters that are designed with three scales of center frequencies, $\sqrt{2}/2$, $\sqrt{2}$, and $b\sqrt{2}/4 = 2\sqrt{2}$ when $b = 8$, as well as eight orientations of angles, 0, 30, 60, 90, 120, 150, 180 and 210 degrees.

4 Segmentation Procedures

For each pixel i, the collection of training set is denoted as $\{\mathcal{Y}^{(i)}, \mathcal{X}^{(i)}\}$, where $\mathcal{X}^{(i)}$ is defined in equation (4) and $\mathcal{Y}^{(i)} = \{y_l^{(i)}(t_j), j = 1, \cdots, m, l \in \mathcal{N}_q^{(i)}\}$, $i = 1, \cdots, n$. The model of DSIR for this training set becomes

$$y_l^{(i)}(t_j) = f^{(i)}(\beta_1^{(i)\prime}\mathbf{x}_l^{(i)}(t_j), \cdots, \beta_K^{(i)\prime}\mathbf{x}_l^{(i)}(t_j), \epsilon_l^{(i)}(t_j)), \tag{5}$$

where $i = 1, \cdots, n$, $j = 1, \cdots, m$, and $l \in \mathcal{N}_q^{(i)}$.

We will describe the algorithm for estimate the *e.d.r.* directions for each pixel i based on the training data as follows.

1. Compute sample mean and sample covariance matrix of $\mathbf{x}_i(t_j)$'s:

$$\begin{aligned}\bar{\mathbf{x}}^{(i)} &= (m|\mathcal{N}_q^{(i)}|)^{-1} \sum_{l \in \mathcal{N}_q^{(i)}} \sum_{j=1}^{m} \mathbf{x}_l^{(i)}(t_j), \\ \hat{\Sigma}_{\mathrm{XX}}^{(i)} &= \left(m|\mathcal{N}_q^{(i)}| - 1\right)^{-1} \sum_{l \in \mathcal{N}_q^{(i)}} \sum_{j=1}^{m} [\mathbf{x}_l^{(i)}(t_j) - \bar{\mathbf{x}}^{(i)}][\mathbf{x}_l^{(i)}(t_j) - \bar{\mathbf{x}}^{(i)}]', \qquad (6)\end{aligned}$$

   where $|\mathcal{N}_q^{(i)}|$ is the number of observations in $\mathcal{N}_q^{(i)}$.

2. Sort the data by the range of $\{y_l^{(i)}(t_j), j = 1, \cdots, m,\ l \in \mathcal{N}_q^{(i)}\}$ and divided the data set into $H^{(i)}$ slices. Denoted these slices by $I_1, \cdots, I_{H^{(i)}}$. Let the proportion of all observed $y_l^{(i)}(t)$'s that falls in hth slice be $p_h^{(i)}$.

3. Within each slice, compute the sample mean of $\mathbf{x}_l^{(i)}(t_j)$'s, denoted by $\bar{\mathbf{x}}_h^{(i)}$; that is

$$\bar{\mathbf{x}}_h^{(i)} = (m|\mathcal{N}_q^{(i)}|p_h^{(i)})^{-1} \sum_{\{j,l: y_l^{(i)}(t_j) \in I_h, l \in \mathcal{N}_q^{(i)}, j=1,\cdots,m\}} \mathbf{x}_l^{(i)}(t_j), \quad h = 1, \cdots, H^{(i)}. \qquad (7)$$

4. Principal component analysis is conducted for the data $\bar{\mathbf{x}}_h^{(i)}$, $h = 1, \cdots, H^{(i)}$, with weights as follows. The weighted covariance matrix,

$$\hat{\Sigma}_W^{(i)} = \sum_{h=1}^{H} p_h^{(i)} (\bar{\mathbf{x}}_h^{(i)} - \bar{\mathbf{x}}^{(i)})(\bar{\mathbf{x}}_h^{(i)} - \bar{\mathbf{x}}^{(i)})',$$

   is computed first. Then, the eigenvalues and eigenvectors for $\hat{\Sigma}_W^{(i)}$ with respect to $\hat{\Sigma}_{\mathrm{XX}}^{(i)}$ are found by solving

$$\hat{\Sigma}_W^{(i)} \hat{\beta}_k^{(i)} = \hat{\lambda}_k^{(i)} \hat{\Sigma}_{\mathrm{XX}}^{(i)} \hat{\beta}_k^{(i)},$$

   where $k = 1, \cdots, p$ and $\hat{\lambda}_1^{(i)} \geq \hat{\lambda}_2^{(i)} \geq \cdots \geq \hat{\lambda}_p^{(i)}$.

5. Then, these first K eigenvectors $\hat{\beta}_k^{(i)}$'s are used as the projection directions for the pixel i.

We denoted the set $\{\hat{\beta}_k^{(i)}, k = 1, \cdots, K\}$ as DSIR projection directions for the pixel i. The segmentation method described above is applied to each pixel of an image respectively.

The algorithm demands less computational cost because the major computation cost is spent at the eigensystem decomposition.

For supervised segmentation, the class labels in the training set belong to G classes. That is,

$$y_l^{(i)}(t_j) \in \{1, 2, \cdots, G\},$$

where $j = 1, \cdots, m,\ l \in \mathcal{N}_q^{(i)}, i = 1, \cdots, n$. Those $p \times 1$ feature vectors, $\mathbf{x}_l^{(i)}(t_j)$, can be obtained from the space, frequency, or space-frequency domain. The DSIR algorithm can be applied to obtain the *e.d.r.* directions, $\hat{\beta}_1^{(i)}, \cdots, \hat{\beta}_K^{(i)}$, where $K \leq \min\{G-1, p\}$. We can then project those feature vectors onto these *e.d.r.* directions. The means of the projected feature vectors for each classes in the training set are the centroids. A pixel i in a test (or training) image is classified (or predicted) into the gth class if the projected feature vector of that pixel is closest to the centroid of gth class. Other classification techniques can be applied as well, like nonparametric discrimination by kernel density estimation, classification trees, and so on. Further discussions regarding the determination of K are reported in Li (1991, 2000) and Ferre (1998).

5 Experimental Results

The above segmentation procedures are performed on simulated and MR image sequences. All images used have 256 grey values and there is only one moving object. Thus, there are only 2 classes, object and background. The first projection direction by DSIR is used in segmentation. The training set is formed by the first three image sequences and the test set consists the next three image sequence. Finally, comparison studies of principal component analysis (PCA), independent component analysis (ICA), SIR, and DSIR are conducted for investigation.

5.1 Simulation Studies

Six successive frames of images with a resolution of 80×80 pixels are simulated. Figure 1 displays the image sequence. The object looks like a ring and deforms from southwest to northeast, that is, from left to right and top to bottom. They are indexed by frame 1 to 6. There are two classes in each time frame, object and background, with gray levels at 90 and 120 respectively. The added noise are randomly distributed from a Gaussian distribution with mean 0 and standard deviation 20. The block size is chosen as 4×4. Hence the feature vectors have 16 variates in the space domain and 10 variates in the frequency domain. The features in the space-frequency domain are constructed by choosing a block size with 8×8 and a bank of $r = 3 \times 8 = 24$ Gabor filters. Therefore the dimension of the feature vector is 48 in the space-frequency domain. We first use frames 1 to 3 as training images and the frame 4 as the test image. After supervised segmentation, the predicted class labels for frame 4 are obtained. The predicted class labels for frame 4 are further added in the training set to predict frame 5 and 6. We then performed the same steps by choosing the frame 1 to 4 as training images and predict the frame 5. Again, the predicted class labels for frame 5 are further added to predict frame 6. Segmentation results in three feature domains are shown in Figure 2, 3, and 4. Table 1 and 2 report the classification and prediction error rates for six frames in the space, frequency and space-frequency domains.

¿From these tables, the accuracy for predicting the object in frame 4 to 6 reaches over 97 percent. Because this object is of the same gray level with added random noises, the features in the space domain work well in the classification and prediction of the first 5 frames because the average of them can distinguish the object and background. However, the prediction error of frame 6 increases due to the motion that will mask the distinction between projected features between object and background in the space domain as one can see in Figure 2. The features in the frequency domain does not work as well as those in the space domain due to the motion artifacts as one can see in Figure 3. The features in the space-frequency domain also have the least motion artifacts in Figure 4. Although the classification and prediction errors in the space-frequency domain are bigger, the segmented boundaries of the object for frame 5 and 6 are more smooth. Thus, for simple textures in these simulation studies, the features in the space domain are sufficient for DSIR in classification and prediction. For complicated textures in the section of comparison studies, the spatial-frequential analysis is needed for DSIR in classification and prediction.

5.2 Clinical Studies of MR Image Sequences

The technique of MR imaging has become a widely used modality as a diagnosis tool for its highly-quality contrast of soft tissue, non-invasiveness and functionalities. It is needed to identify components corresponding to different types of tissues and structure with computers automatically. For example, human brain tissues can be classified into three types by MR images, including gray matter, white matter, and cerebro-spinal fluid (CSF). We investigate the performance of our new method on a MR image sequence of the epi- and endo-cardial surfaces of myocardium. Since the heart is an organ that exhibits motion, the examination of its image characteristics with 2D images sequence reveals useful information about its physical condition. Figure 5 shows one of the images of a heart MR image sequence, where the rectangle box is the region-of-interest (ROI). The image sequence of 6 time frames of ROI is shown in Figure 6. The goal in this experiment is to extract the boundary description of the inner and outer walls for the endocardium of the left ventricle. The training set can be obtained from first three images by medical expert. If we apply the SIR method as a supervised segmentation technique to segment a single image and predict the next image, the boundary of the endocardium is not segmented clearly in Figure 7. The boundary of the endocardium is connected with other tissue both in the classification result of the left image and the prediction result of the right image. Hence, it is necessary to apply the DSIR model to segment the boundary of the endocardium from a sequence of MR images. Use our proposed procedure, the segmentation results in three feature domains are shown in Figure 8, 9, and 10. The errors in comparisons to the ideal boundaries by the doctor are reported in Table 3 and 4. The new approach of DSIR can have successful segmentation results for this sequence of MR images.

5.3 Comparison Studies

To compare the performances of projection directions found by different approaches, we first apply the PCA, ICA, and SIR on a single image. Then, we will extend the study to motion sequences by applying DSIR.

For the image segmentation of a single image, image patches of projection directions found by PCA or ICA usually resemble edges with different lengths and widths that capture the spatial information of position, orientations, spatial frequencies and phases of objects presented in the images (Bell and Sejnowski, 1997, Lewicki and Olshausen, 1999, Lee and Lewicki, 2002). Typically, one image is used and a subset of images patches are randomly selected to perform the PCA or ICA. Then, the leading PCA/ICA directions represent the features of objects in the images, such as lips in face recognition.

For example, we will consider the first image from two consecutive images with natural textures in Figure 11 (a). Every image is of size 80 × 80 pixels that has a object of one texture on the background with another texture. The training set of classification is obtained by randomly selecting 1000 image patches with size 8×8 in the first image and the second image is used as the test set for prediction. The PCA and ICA components are displayed as image patches by the decreasing order in Figure 12 (a) and Figure 13 (a). The PCA components show an increasing order of spatial frequency. We have applied the algorithm of FastICA in Hyvarinen and Oja (2000) to estimate the ICA components one by one. That is, the deflation scheme is used to find the first ICA component by maximizing a contrast function, the second ICA component by maximizing the contrast function with a decorrelation procedure with respect to the first ICA component, and so forth (Hyvarinen 1999). The ICA components also have an increasing order of spatial frequency. However, the PCA and ICA components are different because they are derived by different criteria. For the PCA/ICA components with the high order of frequencies, the image patches have the checkerboard-like patterns locally. The checkerboard effect is produced when the resulting directions are redundant (Lewicki and Olshausen, 1999, Bins and Draper, 2001).

Similarly to the PCA/ICA, we can also apply the SIR on the first image in Figure 11 (a). There are two classes. Unlike the PCA/ICA approaches, the information of class labels in the training set is used in the SIR. The resulting projection directions are shown in decreasing order of the corresponding non-negative eigenvalues in Figure 14 (a). The first eigenvalues λ_1 corresponding to first projection direction accounts for $\lambda_1/\sum_i \lambda_i = 99.9\%$ of the variance. Although the first eigenvector has checkerboard patterns that appear in some local spots, the weights in the middle region have more contribution to classify class labels. The other image patches have more evident checkerboard-like patterns. Since there are two classes in this example, the first eigenvector of SIR is the only component that has the most information for classification in the perspective of Fisher's linear discriminant analysis (Chen and Li, 2001).

Figure 17 (a) shows the segmentation results by using the first projection direction in the space domain. Those 1000 image patches randomly selected in the first image and the corresponding class labels are used as the training set to classify and predict the labels of all pixels in the first and second image. So the segmentation in the first image displays the classification results of those 1000 selected pixels and the prediction results of the remaining pixels. The segmentation in the second image shows the prediction results. Indeed, the SIR can be used to classify and predict the location of these two textures.

Figure 15 (a) and 16 (a) show the segmentation results by using the first PCA and

ICA components in the space domain, respectively. The classification and prediction error rates of PCA, ICA and SIR are reported in Table 5. These results show that PCA, ICA and SIR can identify the object in the space domain with averaged error rates lower than 0.0352. Moreover, the SIR performed better than PCA and ICA with classification error rate 0.0340 and prediction error rate 0.0289.

Next, we re-scale the grey levels of the objects so that the average gray levels of local blocks in the object texture and background texture are similar as in Figure 11 (b). This makes the task of segmentation more difficult. Use the same setting as the previous example, we apply the PCA, ICA, and SIR to the first image in Figure 11 (b). The PCA and ICA components are displayed in Figure 12 (b) and Figure 13 (b). The patterns are similar to those found in the previous example. The SIR components are shown in Figure 14 (b). The first eigenvalues corresponding to first projection direction also accounts for 99.9% of the variance. The remaining image patches show more checkerboard-like patterns than the previous example.

Figure 15 (b), 16 (b) and 17 (b) display the classification (at the left column) and prediction results (at the right column) in the space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row) for PCA, ICA and SIR. According to the results, the PCA, ICA and SIR fail to identify the object in these three feature domains. The averaged classification and prediction error rates range from 0.2429 to 0.4444. This failure can be attributed to the fact that the training set are formed globally. The change of local structures can affect the classification and prediction results. This motivates us to develop the DSIR approach on motion segmentation, a new and effective method to explore stable features in local blocks by spatial, frequential and temporal analysis.

To investigate the capability of DSIR in capturing the directions of features localized in space and time, we will consider four consecutive frames of natural textures images given in Figure 18. Each image frame is of size 80 × 80 pixels that has a moving object of one texture on the background with another texture. The object can deform shape and change location, which make the structure modelling difficult. The first three frames are used as the training set to construct projection directions and the last image frame is used as the test set.

In order to investigate the patterns of projection directions, the feature vectors are formed by choosing a block size 8 × 8 which resulting 64 variates in the space domain, $(8\times 8)/2 + 2 = 34$ variates in the frequency domain, and $2\gamma = 48$ variates in the space-frequency domain. For each pixel i in the image, those 64 gray levels (e.g. in the space domain) are averaged (or projected) by the weights in the projection direction of β. The class label is assumed to be determined by a nonlinear function through the projected values and random noises. Assume the weights in projection directions hold for the next time frame, then the DSIR model in equation (2) fits in this situation. We then choose the 9-order neighborhood system $|\mathcal{N}_9^{(i)}| = 49$ so that the number of observations becomes $3 \times 49 = 147$, which is bigger than the number of parameters in projection directions of β's, like 64, 34, and 48 in those three domains. In our experiments, the selection of block size and neighborhood system is determined so that the number of observation is bigger than the number of observations and the classification error in the training set is reduced mostly.

We selected two rectangles of size 8 × 8 (labelled as A and B) to investigate the first projection directions by DSIR. These two rectangles contain a part of background and a part of moving textures. The resulting patterns are shown in Figure 19, in which each image patch represents the first projection direction by DSIR and corresponds to a single pixel in the rectangle A or B. The brightness of each image pixel is proportional to its weight in the projection direction. Figure 20 shows the segmentation results in the space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row). The first three images in each domain show the classification results and the last image displays the prediction result. These segmentation results demonstrate that the moving boundaries of objects are identified by the proposed approaches in all three feature domains with small error rates in Table 7 and 9 (a).

Next, we apply DSIR on the re-scaled images sequences given in Figure 21, in which the average gray levels of local blocks in the moving texture and background texture are similar. This makes the tasks of motion segmentation more challenging. Use the same setting as above, Figure 22 display the projection directions for two rectangles, A and B. The results are more spatial varying than those in Figure 19. Figure 23 shows the segmentation results. The classification and prediction error rates are reported in Table 8 and 9 (b), which are reduced significantly comparing to those by PCA, ICA and SIR approaches. The performance in the space domain is worse than that in the frequency and space-frequency domains. This illustrates the necessity of DSIR with spatial-frequential analysis for motion segmentation.

Unlike the PCA, ICA, and SIR approaches on a single image, the DSIR uses local blocks from the successive images to construct the projection directions. In our experimental results, there are no apparent patterns of DSIR components like those of PCA/ICA components. The image patches of DSIR components contain the spatiotemporal information regarding the weights at pixels that could classify the class labels mostly, which are different from the spatial information contained in the PCA/ICA components.

6 Conclusion and Discussion

We have extended the model of SIR to the dynamic model of DSIR. Integrating with the spatial-frequential analysis, it is useful to the supervised segmentation of image sequences with complicated textures and deformation. The projection directions learned in the training set can be used to predict the next image in the test set. This prediction can be further added in the training set for the prediction of another new image that comes in time order. The advantage for this recursive procedure of updating is that we can adopt the results of segmentation in previous images to predict future images. We only need a small number of sequence of images in the training set by our method with the capability of feature extraction and dimension reduction. The prediction errors will accumulate as the segmentation process goes. Further tune-ups by image processing techniques, such as smoothing, erosion, dilation, and so on, can be applied to obtain preferred segmentation.

Given a sequence of images of size m, the selection of block size (b), the order of neighborhood system (q), the parameters in the feature extraction (like the number of Gabor filters, r, in the space-frequency domain) and other parameters need further studies. The application of other classification tools and dimension reduction techniques, like the principal hessian directions algorithm in Li (1992), are also of interest in future studies. Besides the application of motion segmentation, DSIR can provide a useful tool for mining the relationship of responses and regressors over time from the perspective of dimension reduction. It will be very interesting to explore other applications that DSIR can be useful.

Acknowledgements

The authors thank Dr. Hui-Cheng Cheng at Veterans General Hospital, Taipei, Taiwan, for providing and explaining the clinical trial data analyzed herein. The authors also thank the associate editor and referees for their valuable comments. This research was supported by the grants from National Science Council at Taiwan, R. O. C.

Appendix: Proof of Theorem 1

Assume that $E[\mathbf{x}(t)] = 0$ without loss of generality. Let $B = (\beta_1, \cdots, \beta_K)$, which is a p by K matrix.

$$\begin{aligned} E[\mathbf{x}(t)|\beta_1'\mathbf{x}(t), \cdots, \beta_K'\mathbf{x}(t)] &= E[\mathbf{x}(t)|B'\mathbf{x}(t)] \\ &= [cov(B'\mathbf{x}(t))^{-1} cov(B'\mathbf{x}(t), \mathbf{x}(t))]' B'\mathbf{x}(t) \qquad (1) \end{aligned}$$

$$= \quad [(B'cov(\mathbf{x}(t))B)^{-1}B'cov(\mathbf{x}(t))]'B'\mathbf{x}(t)$$

The equation (1) is obtained by applying multiple linear regression of each component of $\mathbf{x}(t)$ separately against $B'\mathbf{x}(t)$. Let $\mathbf{k}(y) = (B'cov(\mathbf{x}(t))B)^{-1}E[B'\mathbf{x}(t)|y(t)]$, then

$$\begin{aligned} E[\mathbf{x}(t)|y(t)] &= E(E[\mathbf{x}(t)|B'\mathbf{x}(t), \epsilon(t)]|y(t)) \\ &= E(E[\mathbf{x}(t)|B'\mathbf{x}(t)]|y(t)) \\ &= (B'cov(\mathbf{x}(t)))\mathbf{k}(y). \end{aligned}$$

Appendix 2: Supporting Figures and Tables

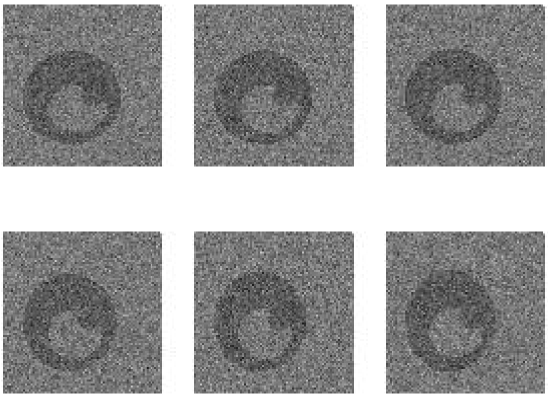


Figure 1: The simulated image sequence with the size of 80 × 80 pixels. The object looks like a ring and deforms. The order of the sequence is from left to right and from top to bottom.

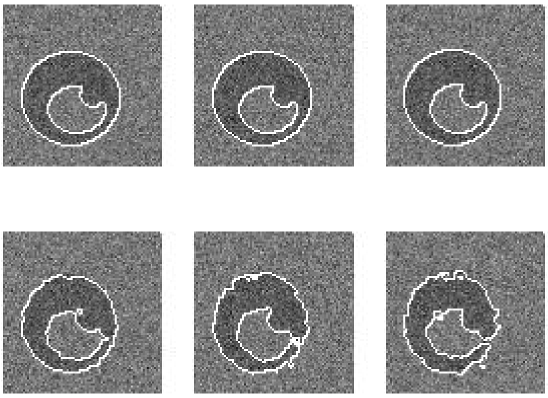


Figure 2: The segmentation results are displayed for the simulated images in the space domain. The top three images display the classification results and the bottom three images show the prediction results.

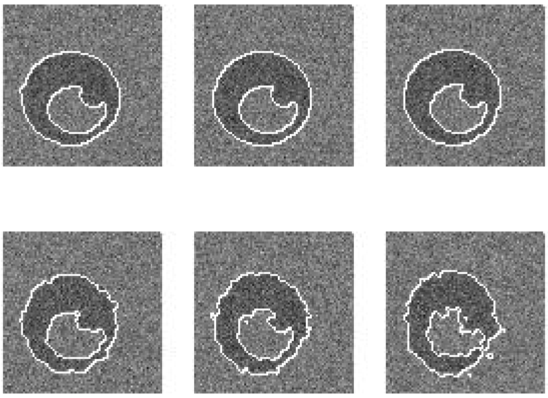


Figure 3: The segmentation results are displayed for the simulated images in the frequency domain. The top three images display the classification results and the bottom three images show the prediction results.

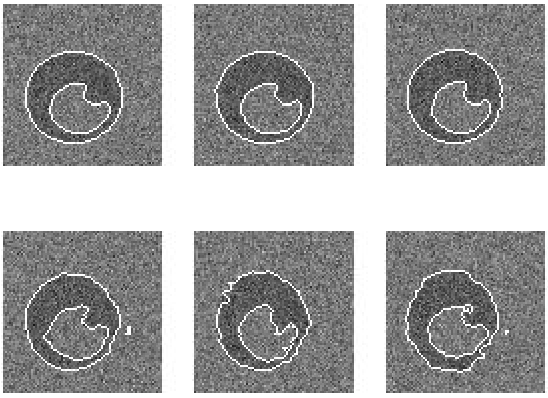


Figure 4: The segmentation results are displayed for the simulated images in the space-frequency domain. The top three images display the classification results and the bottom three images show the prediction results.

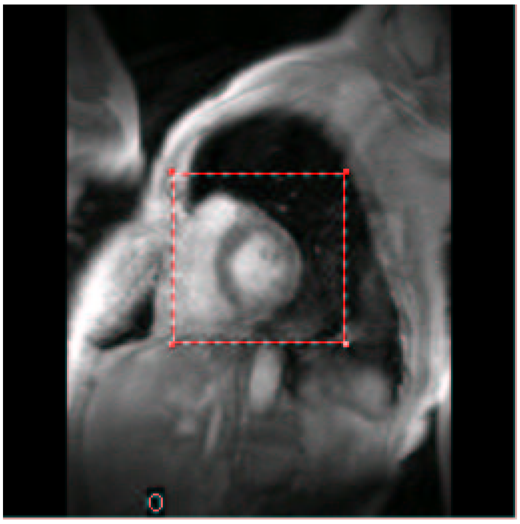


Figure 5: A MR image of human heart in time frame 1, where the rectangle box is the region-of-interest (ROI).

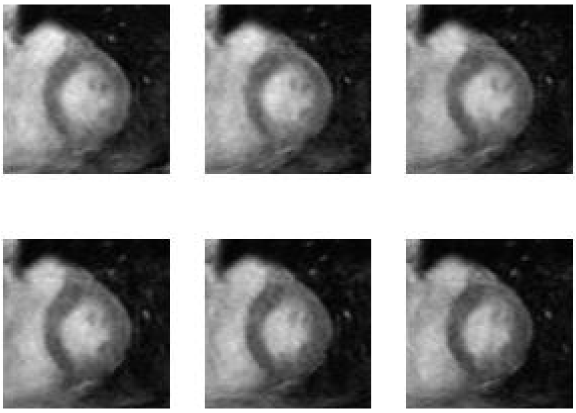


Figure 6: The MR image sequence of myocardium with the size of 84 × 84 pixels in the ROI. The frame order is from left to right and from top to bottom.

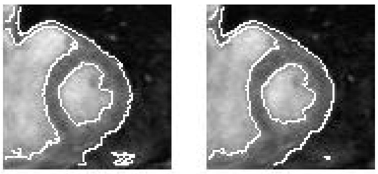


Figure 7: The segmentation result of applying SIR in the supervised segmentation of myocardium. The left image is obtained by the classification of frame 1 and the right image is obtained by the prediction of frame 2.

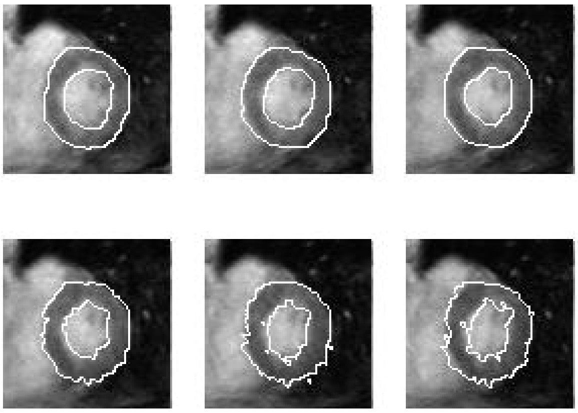


Figure 8: The segmentation results are displayed for the MR image sequence of myocardium in space domain. The top three images display the classification results and the bottom three images show the prediction results.

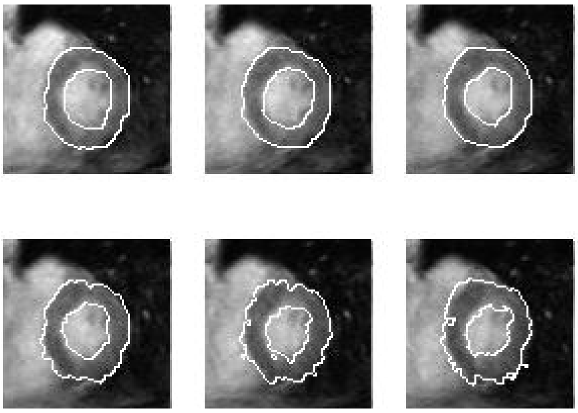


Figure 9: The segmentation results are displayed for the MR image sequence of myocardium in frequency domain. The top three images display the classification results and the bottom three images show the prediction results.

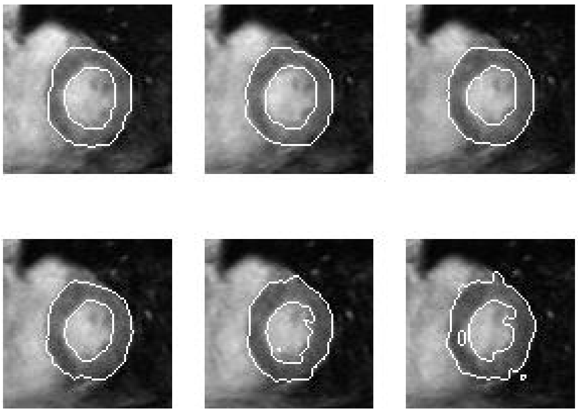


Figure 10: The segmentation results are displayed for the MR image sequence of myocardium in space-frequency domain. The top three images display the classification results and the bottom three images show the prediction results.

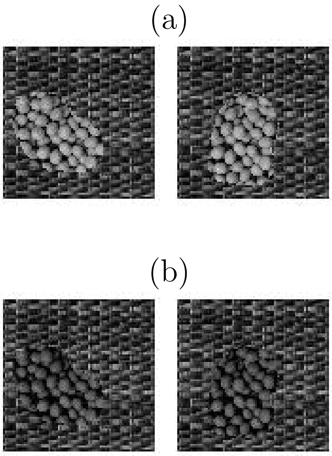
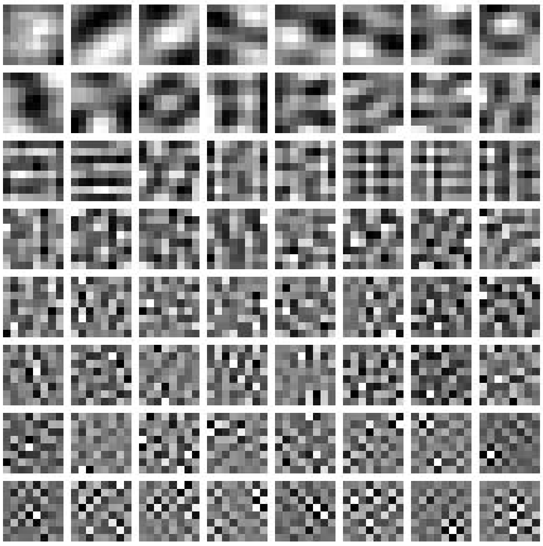


Figure 11: (a) Two consecutive frames of a moving object with one texture and a background with another texture. (b) The texture image sequence is re-scaled so that the gray levels of the local blocks in object and background have similar mean values.

(a)



(b)

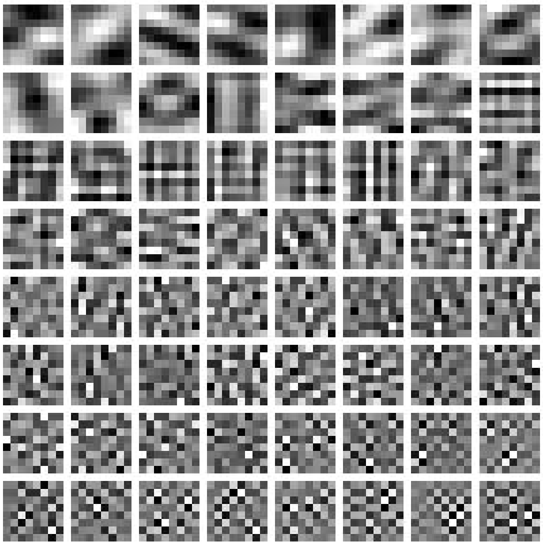


Figure 12: The PCA components are displayed in the decreasing order of non-negative eigenvalues, from left to right and top to bottom, where (a) is for the first image in Figure 11 (a) and (b) is for the first image in Figure 11 (b).

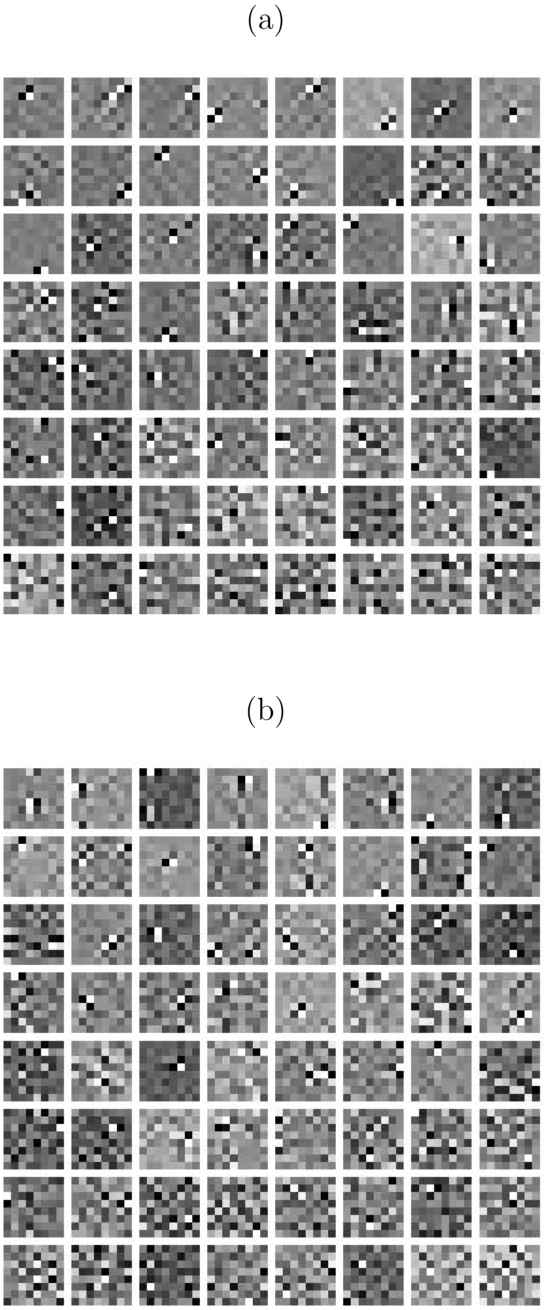


Figure 13: The ICA components using a deflation scheme are displayed, where (a) is for the first image in Figure 11 (a) and (b) is for the first image in Figure 11 (b).

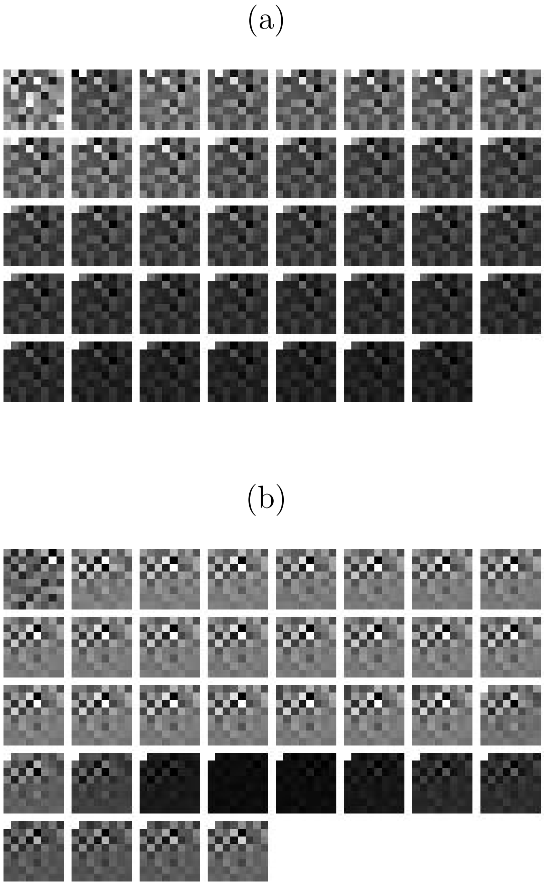


Figure 14: The projection directions by SIR are shown in the decreasing order of non-negative eigenvalues, from left to right and top to bottom, where (a) is for the first image in Figure 11 (a) and (b) is for the first image in Figure 11 (b).

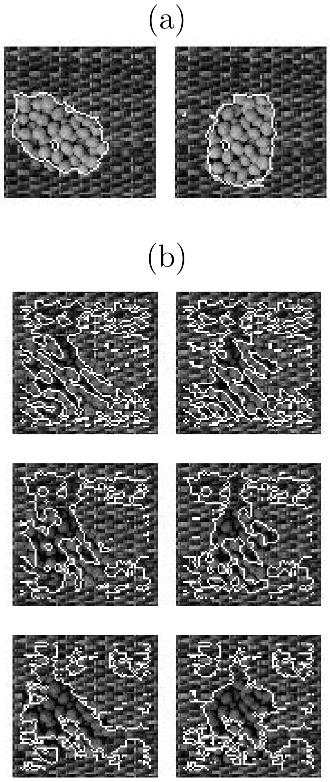


Figure 15: The segmentation results of applying PCA on a single image are displayed. The images at left column are obtained by the classification and the images at right column are obtained by the prediction. (a) is for the first two image in Figure 11 (a) in space domain and (b) is for the first two image in Figure 11 (b) in space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row).

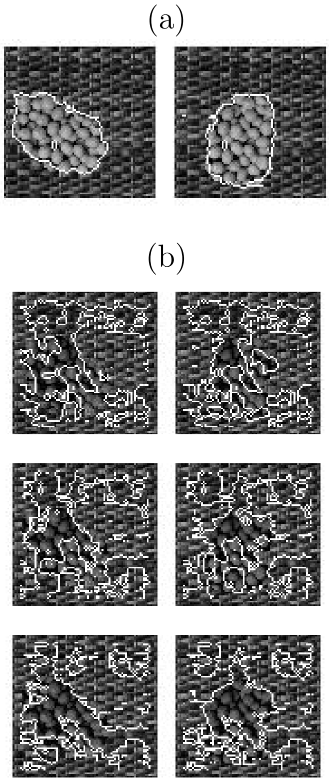


Figure 16: The segmentation results of applying ICA on a single image are displayed. The images at left column are obtained by the classification and the images at right column are obtained by the prediction. (a) is for the first two image in Figure 11 (a) in space domain and (b) is for the first two image in Figure 11 (b) in space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row).

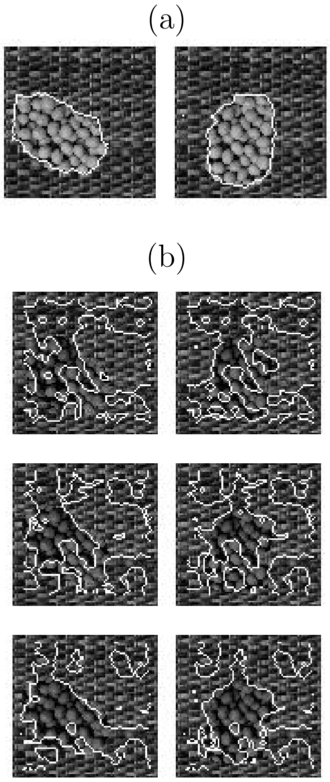


Figure 17: The segmentation results of applying SIR on a single image are displayed. The images at left column are obtained by the classification and the images at right column are obtained by the prediction. (a) is for the first two image in Figure 11 (a) in space domain and (b) is for the first two image in Figure 11 (b) in space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row).

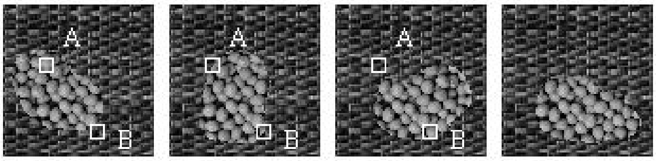


Figure 18: Four consecutive frames of a moving and deforming object with one texture in a background with another texture. Two rectangles with size 8 × 8 (labelled as A and B) are selected to display the first DSIR projection directions.

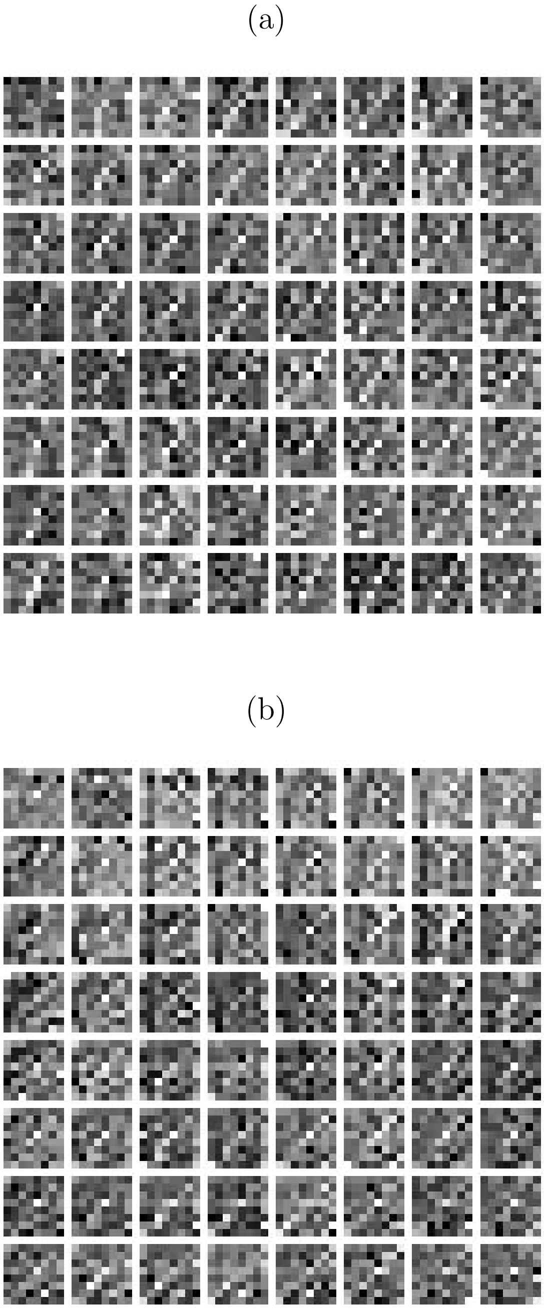


Figure 19: The first projection directions by DSIR for the selected blocks A (a) and B (b) in Figure 18 are displayed. Each image patch represents the first projection direction by DSIR on a single pixel in the rectangles A or B.

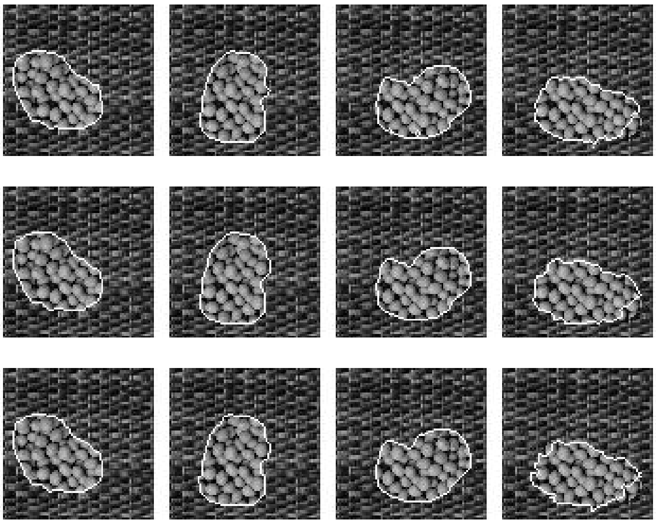


Figure 20: The segmentation results of Figure 18 are displayed for the texture images sequence in the space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row). For each domain, the first three images show the classification results and the last one displays the prediction result.

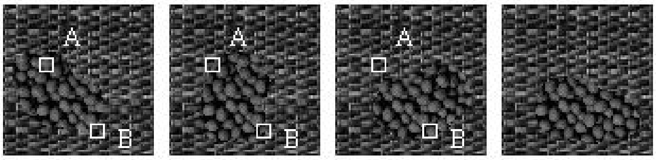


Figure 21: The texture image sequence is re-scaled from Figure 18 so that the gray levels of the local blocks in object and background have similar mean values.

(a)

(b)

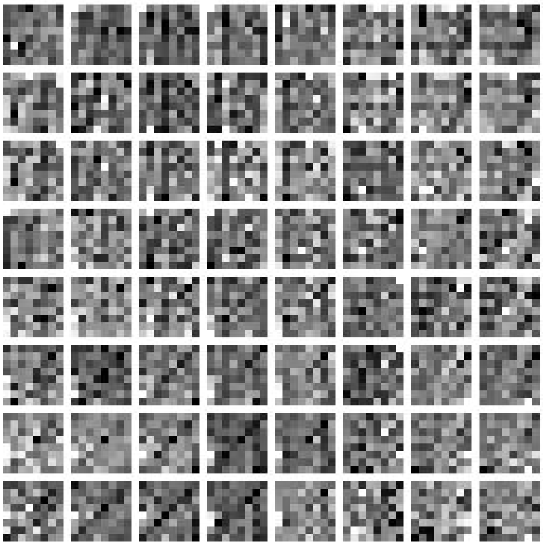


Figure 22: The first projection directions by DSIR for selected blocks A (a) and B (b) in Figure 21 are displayed. Each image patch represents the first projection direction by DSIR on a single pixel in the rectangles A or B.

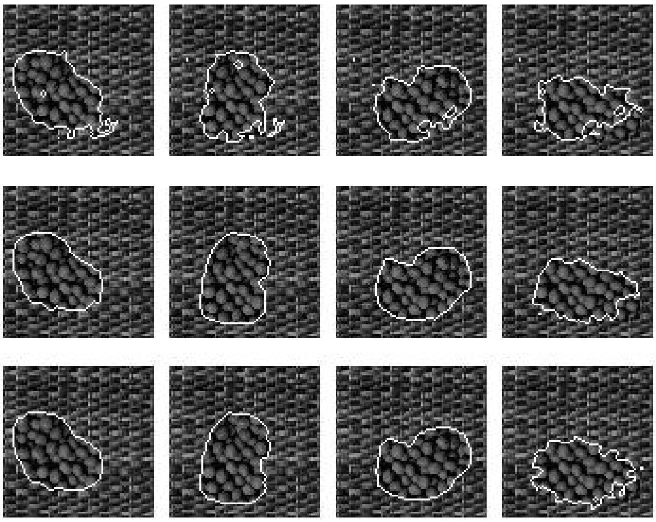


Figure 23: The segmentation results of Figure 21 are displayed for the texture images sequence in the space domain (at the top row), the frequency domain (at the second row) and the space-frequency domain (at the bottom row). For each domain, the first three images show the classification results and the last one displays the prediction result.

Table 1: The classification error rates for the simulated image sequence, where Obj represents the object, Bg denotes the background, and Total means the average error of the whole image.

	Frame 1			Frame 2			Frame 3		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0015	0.0004	0.0007	0.0015	0.0004	0.0007	0.0000	0.0004	0.0003
FFT	0.0127	0.0033	0.0054	0.0083	0.0021	0.0035	0.0089	0.0037	0.0049
Gabor	0.0045	0.0036	0.0038	0.0090	0.0036	0.0050	0.0082	0.0016	0.0033

Table 2: The prediction error rates for the simulated image sequence, where Obj represents the object, Bg denotes the background, and Total means the average error of the whole image.

	Frame 4			Frame 5			Frame 6		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0459	0.0104	0.0184	0.0337	0.0096	0.0150	0.0572	0.0167	0.0258
FFT	0.0452	0.0152	0.0219	0.0539	0.0171	0.0254	0.0429	0.0241	0.0283
Gabor	0.0505	0.0122	0.0220	0.0494	0.0169	0.0253	0.0474	0.0201	0.0276

Table 3: The classification error rates for the MR image sequence, where Obj represents the object, Bg denotes the background, and Total means the average error of the whole image.

	Frame 1			Frame 2			Frame 3		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0132	0.0013	0.0034	0.0068	0.0027	0.0035	0.0032	0.0016	0.0019
FFT	0.0167	0.0023	0.0048	0.0188	0.0044	0.0070	0.0097	0.0033	0.0045
Gabor	0.0221	0.0032	0.0069	0.0154	0.0043	0.0066	0.0081	0.0040	0.0048

Table 4: The prediction error rates for the MR image sequence, where Obj represents the object, Bg denotes the background, and Total means the average error of the whole image.

	Frame 4			Frame 5			Frame 6		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0719	0.0216	0.0310	0.0630	0.0313	0.0370	0.0964	0.0319	0.0440
FFT	0.0832	0.0191	0.0312	0.0750	0.0302	0.0382	0.1264	0.0399	0.0562
Gabor	0.0816	0.0145	0.0289	0.0503	0.0337	0.0370	0.0899	0.0398	0.0505

Table 5: The classification and prediction error rates of PCA, ICA and SIR for two consecutive texture images in Figure 11 (a) in the space domain.

	Classification			Prediction		
Methods	Obj	Bg	Total	Obj	Bg	Total
PCA	0.0608	0.0339	0.0413	0.0665	0.0220	0.0343
ICA	0.0519	0.0336	0.0387	0.0624	0.0233	0.0342
SIR	0.0416	0.0311	0.0340	0.0536	0.0195	0.0289

Table 6: The classification and prediction error rates of PCA, ICA and SIR for two consecutive texture image in Figure 11 (b).

		Classification			Prediction		
Methods	Feature	Obj	Bg	Total	Obj	Bg	Total
PCA	Space	0.4348	0.4480	0.4444	0.4389	0.4457	0.4438
	FFT	0.3652	0.4022	0.3920	0.3514	0.3990	0.3858
	Gabor	0.2130	0.3189	0.2890	0.2673	0.3307	0.3127
ICA	Space	0.3980	0.4371	0.4263	0.3853	0.4340	0.4205
	FFT	0.3754	0.4394	0.4218	0.3921	0.4457	0.4309
	Gabor	0.2129	0.3189	0.2889	0.2673	0.3307	0.3127
SIR	Space	0.4908	0.3483	0.3875	0.3826	0.4358	0.4211
	FFT	0.3530	0.4205	0.4020	0.3751	0.4280	0.4134
	Gabor	0.1570	0.2767	0.2429	0.2028	0.2914	0.2662

Table 7: The classification error rates of DSIR for texture image sequence in Figure 18.

	Frame 1			Frame 2			Frame 3		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0034	0.0041	0.0039	0.0122	0.0018	0.0047	0.0134	0.0016	0.0049
FFT	0.0089	0.0031	0.0047	0.0047	0.0023	0.0030	0.0087	0.0044	0.0056
Gabor	0.0055	0.0027	0.0035	0.0088	0.0040	0.0054	0.0061	0.0041	0.0046

Table 8: The classification error rates of DSIR for texture image sequence in Figure 21.

	Frame 1			Frame 2			Frame 3		
Feature	Obj	Bg	Total	Obj	Bg	Total	Obj	Bg	Total
Space	0.0177	0.0355	0.0306	0.0909	0.0184	0.0385	0.0599	0.0177	0.0295
FFT	0.0137	0.0049	0.0073	0.0095	0.0042	0.0056	0.0061	0.0073	0.0069
Gabor	0.0034	0.0046	0.0042	0.0088	0.0051	0.0062	0.0067	0.0062	0.0064

Table 9: The prediction error rates of DSIR for Frame 4 in Figure 18 and 21.

	(a) Figure 18			(b) Figure 21		
Feature	Obj	Bg	Total	Obj	Bg	Total
Space	0.0655	0.0161	0.0298	0.2438	0.0265	0.0869
FFT	0.0858	0.0247	0.0417	0.1458	0.0161	0.0522
Gabor	0.0932	0.0232	0.0432	0.1262	0.0259	0.0545

References

Bell, A. J. and Sejnowski, T. J. (1997). The ’independent components’ of natural scenes are edge filters. *Vision Research* **37**(23), 3327-3338.

Bins, J. and Draper, B. A. (2001). Feature selection from huge feature sets. *International Conference on Computer Vision* **2**, 159-165.

Bouthemy, P. (1989). A maximum likelihood framework for determining moving edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, 499-511.

Chen, C. H. and Li, K. C. (1998). Can SIR be as popular as multiple linear regression? *Statistica Sinica* **8**, 289-316.

Chen, C. H. and Li, K. C. (2001). Generalization of Fisher’s linear discriminant analysis via the approach of sliced inverse regression. *Journal of the Korean Statistical Society* **30**(2), 193-217.

Chen, C. M., Lu, H. H.-S. and Lin, Y. C. (1999). An early vision based snake model for ultrasound image segmentation. *Ultrasound in Medicine and Biology* **26**, 273-285.

Chen, C. M., Lu, H. H.-S. and Han, K. C. (2001). A textural approach based on Gabor functions for texture edge detection in ultrasound images. *Ultrasound in Medicine and Biology* **27**, 513-534.

Deng, Y. and Manjunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 800–810.

Duncan, J. S., Owen, R. L., Staib, L. H. and Anandan, P. (1991). Measurement of non-rigid motion using contour shape descriptors. *Computer Vision and Pattern Recognition* **91**, 318-324.

Dunn, D., Higgins, W. E. and Wakeley, J. (1994). Texture segmentation using 2-D Gabor elementary functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**, 130-149.

Ferre, L. (1998). Determining dimension in sliced inverse regression and related methods. *Journal of the American Statistical Association* **93**, 132-140.

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks* **10**(3), 626-634.

Hyvarinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Networks* **13**(4-5), 411-430.

Isard, M. and Blake, A. (1996). Contour tracking by stochastic propagation of conditional density. *European Conference on Computer Vision* (1), 343-356.

Lee, T. W. and Lewicki, M.S. (2002). Unsupervised image classification, segmentation, and enhancement using ICA mixture models. *IEEE Transactions on Image Processing* **11**(3), 270-279.

Lewicki, M. S. and Olshausen, B. A. (1999). A probabilistic framework for adaptation and comparison of image codes. *Journal of the Optical Society of America* **16**, 1587-1600.

Li, K. C. (1991). Sliced inverse regression for dimensional reduction (with discussion). *Journal of the American Statistical Association* **86**, 316-342.

Li, K. C. (1992). On principal hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *Journal of the American Statistical Association* **87**, 1025-1039.

Li, K. C. (2000). High dimensional data analysis via the SIR/PHD approach. Lecture Notes that are available at `http://www.stat.ucla.edu/~kcli/`.

Li, K.C., Aragon, Y. and Thomos-Agan, C. (1995). Analysis of multivariate outcome data: SIR and a nonlinear theory of Hotelling's most predictable variates. Technical Report, Department of Statistics, UCLA.

Lu, H. H.-S., Chen, C. M. and Wu, J. S. (2001). Statistical analysis of liver cirrhosis in ultrasound images by fractal dimension, dimension reduction and classification trees. *The 5th World Multi-Conference on Systemics, Cybernetics and Informatics*, Vol. XIII, Part II, 351-356.

Malik, J. and Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America* **A7**, 923-932.

Mikic, I., Krucinski, S. and Thomas, J. D. (1998). Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. *IEEE Transactions on Medical Imaging* **17**, 274-284.

Nguyen, H. T., Worring, M. and Dev, A. (2000). Detection of moving objects in video using a robust motion similarity measure. *IEEE Transactions on Image Processing* **9**, 137 -141.

Soatto, S., Doretto, G. and Wu, Y. (2001). Dynamic textures. *Intl. Conf. on Computer Vision*, 439-446.

Tan, T. N. (1995). Texture edge detection by modelling visual cortical channels. *Pattern Recognition* **28**, 1283-1298.

Vasconcelos, N. and Lippman, A. (2001). Empirical bayesian motion segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **23**, 217 -221.

Weaver, H. J. (1983). Applications of Discrete and Continuous Fourier Analysis. Wiley, New York.

Wu, Y., Zhu, S. C. and Guo, C. (2002). Statistical modelling of texture sketch. *European Conference of Computer Vision.*

Wu, Y., Zhu, S. C. and Liu, X. (2000). Equivalence of Julesz texture ensembles and FRAME models. *International Journal of Computer Vision* **38**, 247-265.

Zhu, S. C., Liu, X. and Wu, Y. (2000). Exploring texture ensembles by efficient Markov Chain Monte Carlo - towards a 'Trichromacy' theory of texture. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 554-569.

Zhu, S. C., Wu, Y. N. and Mumford, D. B. (1998). Filter, random field, and maximum entropy (FRAME): towards a unified theory for texture modelling. *International Journal of Computer Vision* **27**, 107-126.

Institute of Statistics, National Chiao-Tung University, 1001 Ta Hsueh Road, Hsinchu 30050, Taiwan, R.O.C.
E-mail: hmwu.st86g@nctu.edu.tw
E-mail: hslu@stat.nctu.edu.tw