Multisource Image Fusion Method Using Support Value Transform

Sheng Zheng, Wen-Zhong Shi, Jian Liu, Guang-Xi Zhu, and Jin-Wen Tian

***Abstract*—With the development of numerous imaging sensors, many images can be simultaneously pictured by various sensors. However, there are many scenarios where no one sensor can give the complete picture. Image fusion is an important approach to solve this problem and produces a single image which preserves all relevant information from a set of different sensors. In this paper, we proposed a new image fusion method using the support value transform, which uses the support value to represent the salient features of image. This is based on the fact that, in support vector machines (SVMs), the data with larger support values have a physical meaning in the sense that they reveal relative more importance of the data points for contributing to the SVM model. The mapped least squares SVM (mapped LS-SVM) is used to efficiently compute the support values of image. The support value analysis is developed by using a series of multiscale support value filters, which are obtained by filling zeros in the basic support value filter deduced from the mapped LS-SVM to match the resolution of the desired level. Compared with the widely used image fusion methods, such as the Laplacian pyramid, discrete wavelet transform methods, the proposed method is an undecimated transform-based approach. The fusion experiments are undertaken on multisource images. The results demonstrate that the proposed approach is effective and is superior to the conventional image fusion methods in terms of the pertained quantitative fusion evaluation indexes, such as quality of visual information ($Q^{AB/F}$), the mutual information, etc.**

***Index Terms*—Image fusion, mapped least squares support vector machine (mapped LS-SVM), support vector machine (SVM), support value transform (SVT).**

Manuscript received July 5, 2006; revised January 30, 2007. This paper was supported in part by the China National Natural Science Fund (Project No. 60572048), in part by the National 863 Project Fund (Project No. 2003AA135010), in part by Research Grants from the Council of the Hong Kong SAR (Project No. PolyU5153/04E), and in part by the Hubei Province Educational Department Natural Science Fund (Project No. D200613003). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ercan E. Kuruoglu.

S. Zheng is with the Institute of Intelligent Vision and Image Information Best, College of Electrical and Information Engineering, China Three Gorges University, Yichang 443002, China, and also with the Electronic and Information Engineering Department, Huazhong University of Science and Technology (HUST), Wuhan 430074, China (e-mail: zsh@ctgu.edu.cn; zhengsheng6511@sohu.com).

W.-Z. Shi is with the Advanced Research Center for Spatial Information Technology, Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong (e-mail: lswzshi@polyu.edu.hk).

J. Liu is with the Electronic and Information Engineering Department, Huazhong University of Science and Technology (HUST), Wuhan 430074, China (e-mail: bliu@hust.edu.cn).

G.-X. Zhu is with the Chinese Key Laboratory of Optic and Electronics, Huazhong University of Science and Technology (HUST), Wuhan 430074, China (e-mail: gxzhu@hust.edu.cn).

J.-W. Tian is with the State Education Commission Key Laboratory for Image Processing and Intelligent Control, Huazhong University of Science and Technology (HUST), Wuhan 430074, China (e-mail: jwtian@hust.edu.cn).

Digital Object Identifier 10.1109/TIP.2007.896687

I. Introduction

CURRENTLY, numerous imaging sensors have been developed in many fields such as remote sensing, medical imaging, machine vision, and military applications. However, in many cases, one sensor will not give the complete picture of a scenario. For example, many vision-related processing tasks, including edge detection and image segmentation, can be performed more easily when all objects in the scenario are in good focus. However, in practice, this complete picture may not be always feasible since optical lenses of imaging sensor, especially those with long focal lengths, only have a limited depth of field [1]. To effectively integrate the different features from the images pictured using a set of different sensors and produce a single image which preserves all relevant information from the original data set, a classical method is to perform the image fusion. With the availability of multiple image sources, image fusion has emerged as a new and promising research area.

Up until now, many algorithms and software for image fusion have been developed. According to the stage at which the combination mechanism takes place, the existing image fusion methods can be generally grouped into three categories, namely, pixel level, feature level, and decision level. Since both feature-level and decision-level fusion may involve loss of information in the information extraction process, which consequently leads to less accurate fusion results [2], we only focus on pixel-level fusion in this study.

As we know, the goal of the image fusion is to integrate the underlying information of original images and produce a complete picture. The basic assumption of the image fusion is that the underlying information is the salient features of the original images and the salient features are linked to some easily computed information measures, such as the high frequency components of Laplacian pyramid or the coefficients of wavelet transform. To effectively and completely extract the underlying information from the original images, many successful methods have been explored in recent years. They use multiscale transforms as the salient feature extraction tool. These existing methods include Laplacian pyramid (LAP) [3], filter subtract decimate hierarchical pyramid (FSD), contrast pyramid (COP), gradient pyramid (GRP), ratio pyramid (RAP), neural networks [4], and the discrete wavelet transform (DWT) [2], [5]–[7], etc. The problem is the following: May using these simple salient features to fuse the images always produce optimal results? Obviously, the answer is no [8]. Therefore, it is reasonable to assume that if we have a new effective representation of the salient features underlying original images,

the fused results may be further accurate. This is the basic consideration that we have done this study.

It is well known that the support vector machine (SVM) has been developed recently [9] within context of the statistical learning theory and the structural risk minimization. SVM maps input data into a high-dimensional feature space where it may become linearly separable. SVM is a powerful tool for data classification and function estimation, and has been applied to a wide variety of domains such as pattern recognition, object detection, and function estimation, etc. Particularly, the development of a modified version of SVM called least squares SVM (LS-SVM) [10], [11], which resulted in a set of linear equations instead of a quadratic programming problem, extended the application of SVM to on-line applications. The introduction of mapping technique into the LS-SVM extends further the application of SVM to image processing areas [12], such as edge detection [13], interpolation [14], object detection [15], etc. One important point in the SVM is that the data with larger support values can most possibly become the support vectors in the sparse SVM, since the sparse process exploits the fact that the support values have a physical meaning in the sense that they reveal the relative importance of the data points for contributing to the SVM model [16]. This point gives us an important clue, that the support values of the data points may be a measure of the salient feature of the original images if we analyze image under the SVM framework. If we can construct a multiscale support value analysis framework based on the SVM, we may obtain more accurate fusion results than the conventional image fusion methods provide. Based on this consideration, we proposed the support value transform (SVT)-based image fusion method, where the salient features of the original images are represented by their support values.

The rest of this paper is organized as follows. In Section II, we present the support value analysis method. In Section III, the SVT-based image fusion method is described. The image fusion experimental results are provided in Section IV, and the conclusions are drawn in Section V.

II. Support Value Analysis Based on Mapped LS-SVM

A. SVM

Let $\mathbf{x} \in R^d, y \in R, R^d$ represent input space; d is the dimension. By some nonlinear mapping $\phi(\mathbf{x}) : R^d \rightarrow R^q, q$ denotes the dimension of the feature space, $\mathbf{x}$ is mapped into some a prior chosen Hilbert space spanned by the linear combination of a set of functions. The aim of SVM approximation is to estimate the function based on training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$. Vapnik [9] has shown that the basic approximation problem could be solved by finding a set of weights

$$y = f(\mathbf{x}, \boldsymbol{\omega}) = \boldsymbol{\omega}^{\mathrm{T}} \phi(\mathbf{x}) + b \tag{1}$$

where $\boldsymbol{\omega}$ is an element of R^q, such that the following regularized risk function $R(\boldsymbol{\omega})$ is minimized:

$$R(\boldsymbol{\omega}) = \frac{C}{N} \sum_{i=1}^{N} L[y, f(\mathbf{x}, \boldsymbol{\omega})] + \frac{\|\boldsymbol{\omega}\|^2}{2} \tag{2}$$

where $\mathrm{C} > 0$ is constant, $L[y, f(\mathbf{x}, \boldsymbol{\omega})]$ is an ε-insensitive loss function

$$L[y, f(\mathbf{x}, \boldsymbol{\omega})] = \begin{cases} 0, & \text{if } |f(\mathbf{x}, \boldsymbol{\omega}) - y| < \varepsilon \\ |f(\mathbf{x}, \boldsymbol{\omega}) - y| - \varepsilon, & \text{otherwise} \end{cases}. \tag{3}$$

The parameter ε means that the most optimized function $f(\mathbf{x}, \boldsymbol{\omega})$ has at most a deviation of ε from the actual observed target y. By using Lagrange multipliers, the minimization of (2) leads to the dual optimization problem, which will generally lead to a well known solvable convex quadratic programming problem. The resulting approximation estimates are linear.

B. LS-SVM

In LS-SVM [10], [11], SVM has been modified into the following minimization problem:

$$R(\boldsymbol{\omega}) = \frac{\|\boldsymbol{\omega}\|^2}{2} + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^2 \tag{4}$$

subject to equality constraints

$$y_i = \boldsymbol{\omega}^{\mathrm{T}} \phi(\mathbf{x}_i) + b + e_i, \quad i = 1, \ldots, N. \tag{5}$$

The parameter γ is a positive regularization constant. One defines the Lagrangian

$$L = \frac{\|\boldsymbol{\omega}\|^2}{2} + \frac{\gamma}{2} \sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \alpha_i \{\boldsymbol{\omega}^{\mathrm{T}} \phi(\mathbf{x}_i) + b - y_i + e_i\}. \tag{6}$$

The conditions for optimality can be written immediately as the solution to the following set of linear equations after elimination of $\boldsymbol{\omega}, \mathbf{e}$

$$\begin{bmatrix} 0 & \vec{1}^{\mathrm{T}} \\ \vec{1} & \boldsymbol{\Omega} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{Y} \end{bmatrix} \tag{7}$$

where $\boldsymbol{\Omega} = \mathbf{K} + \mathbf{I}/\gamma, \mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j), \mathbf{Y} = [y_1, \ldots, y_N]^{\mathrm{T}}, \vec{1} = [1, \ldots, 1]^{\mathrm{T}}, \boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]^{\mathrm{T}}$.

In LS-SVM, the function f can be approximated by a combination of a set of support vectors

$$f(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{8}$$

where $K(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x})^{\mathrm{T}} \phi(\mathbf{x}_i), i = 1, \ldots, N$ is the kernel function, and α_i is the support value of support vector $\mathbf{x}_i$.

C. Support Values of Mapped LS-SVM

In the 2-D mapped LS-SVM [12], the LS-SVM is used to estimate the underlying intensity surface of an image. In this case, the input vector of the LS-SVM is defined by the pixel coordinate (r, c) and the output is the intensity value $g(r, c)$. For any image block, the input points are usually of the form $\{(r_0 + dr, c_0 + dc): |dr| \leqslant m, |dc| \leqslant n\}$ and all such set of points can

be translated into the same set $\{(dr, dc)\colon |dr| \leqslant m, |dc| \leqslant n\}$ by subtracting (r_0, c_0) from all the vectors, where m and n are, respectively, the half number of the image block's horizontal and vertical pixels. By adding (r_0, c_0) to the set $\{(dr, dc) : |dr| \leqslant m, |dc| \leqslant n\}$, the original image forms are reconstructed. With this mapping technique, the LS-SVM learning and estimating problems can be transformed into the problems with the same set of input vectors but different set of labels. For a specified image block size, the set $\{(dr, dc) : |dr| \leqslant m, |dc| \leqslant n\}$ is an unchanged constant vector space, and the number of input vectors (N) equals $(2m+1)(2n+1)$.

Similar to the standard LS-SVM, the solution of the mapped LS-SVM is to solve a set of linear equations in (7). The explicit solution of (7) is

$$b = \frac{\vec{1}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\mathbf{Y}}{\vec{1}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\vec{1}}, \quad \boldsymbol{\alpha} = \boldsymbol{\Omega}^{-1}(\mathbf{Y} - b\vec{1}). \tag{9}$$

In the defined constant vector space, the input vectors $\{\mathbf{x}_i, i = 1, \ldots, N\}$ are unchanged, and the parameters of the LS-SVM (kernel function K and parameter γ) can be prior chosen, and elements of the vector $\mathbf{Y} = [g_1, \ldots, g_N]^{\mathrm{T}}$ are taken from the $(2m+1) \times (2n+1)$ observed image intensity value matrix. Therefore, the $\boldsymbol{\Omega}$ can be transformed into a constant matrix. In this way, matrices $\mathbf{A}$ and $\mathbf{B}$ can be defined as constants and are precalculated by

$$\mathbf{A} = \boldsymbol{\Omega}^{-1}, \quad \mathbf{B}^{\mathrm{T}} = \frac{\vec{1}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}}{\vec{1}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\vec{1}} \tag{10}$$

then the solution of (9) can be rewritten as

$$b = \mathbf{B}^{\mathrm{T}}\mathbf{Y}, \quad \boldsymbol{\alpha} = \mathbf{A}(\mathbf{Y} - b\vec{1}). \tag{11}$$

These matrices $\mathbf{A}$ and $\mathbf{B}$, however, depend only on the input vectors $\{\mathbf{x}_i, i = 1, \ldots, N\}$, but not on their labels $\{g_i, i = 1, \ldots, N\}$. Therefore, in the mapped LS-SVM, the learning computation complexity can be reduced to only about $O(N^2)$, just only a matrix multiplication. The support value $\{\alpha_i, i = 1, \ldots, N\}$ of the image pixel $\{g_i, i = 1, \ldots, N\}$ in the mapped neighborhood can be easily calculated using (11).

D. Support Value Filter

It is worthy noting that the (11) can be rewritten as follows:

$$\boldsymbol{\alpha} = \mathbf{A}(\mathbf{I} - \vec{1}\mathbf{B}^{\mathrm{T}})\mathbf{Y} = \mathbf{Q}\mathbf{Y} \tag{12}$$

where $\mathbf{Q}$ is a $N \times N$ matrix defined by $\mathbf{A}(\mathbf{I} - \vec{1}\mathbf{B}^{\mathrm{T}})$. The calculation of support value α_i of the observed image pixel g_i in the mapped neighborhood is to multiply the column vector $\mathbf{Y}$ by the ith row vector of matrix $\mathbf{Q}$. That is to say, for the observed pixel g_i in mapped neighborhood, the corresponding support value can be computed individually as a linear combination of the measured intensity values $\{g_i, i = 1, \ldots, N\}$. The weight associated with each g_i is, respectively, determined by the elements of the ith row vector of matrix $\mathbf{Q}$. For a rectangular neighborhood, reshaping the corresponding row vectors of matrix $\mathbf{Q}$, the weight kernels become the support value filters. If the support value of the discrete image in the pixel (r, c) is approximated by the corresponding support value of the input vector in mapped neighborhood center, the support values of the whole image can be computed by convolving the image with the support value filter deduced from the central row vector of matrix $\mathbf{Q}$.

From (7), (10), and (12), we can see that the matrix $\mathbf{Q}$ depends on the input vectors, kernel function K and parameter γ. In the mapped LS-SVM, the input vectors $\{\mathbf{x}_i, i = 1, \ldots, N\}$ are constant for a specified neighborhood size and then the support value filter is only depend on the kernel function K and parameter γ. As the LS-SVM is an optimization method, the number of input vectors should not be too small, and then the size of the mapped neighborhood should be large enough, for example, it can be set to be 5×5 pixels, including 25 input vectors.

The parameter γ controls the solution insensitivity to the error, and more noise requires smaller value of γ. When the γ is infinite, a solution with least squared error is attained. For the computation of support values of an image degraded by more or less noise, the γ should be a little small value, for example, it can be set to be 1 in this study.

Many of the characteristics of the model in (8) are determined by the type of kernel function K. The characterization of a kernel function is done by means of the Mercer's theorem [17]. Every kernel has its advantages and disadvantages. Numerous possibilities of kernels satisfying Mercer's theorem exist. At present, the Gaussian kernel is the most common one due to its good features [18], and this study only focuses on the choice of Gaussian radial basis function (RBF) kernel, $K(\mathbf{x}, \mathbf{x}_i) = \exp(-\|\mathbf{x} - \mathbf{x}_i\|^2/2\sigma^2)$, where the spread parameter σ is closely associated with generalization performance of SVM. In the 2-dimension mapped LS-SVM, the RBF kernel can be rewritten as $\exp(-(|r - r_i|^2 + |c - c_i|^2)/2\sigma^2)$, where (r, c) is the pixel coordinate in the mapped input space. For the neighborhood center, $r = 0$ and $c = 0$.

The computation procedure of the support value filter can be summarized as follows.

1) To give the rectangular neighborhood size of the mapped input vector space, and the parameter γ of the LS-SVM, and the spread parameter σ of Gaussian RBF kernel.
2) To compute the $N \times N$ matrix $\boldsymbol{\Omega}$, where $\boldsymbol{\Omega}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) + \mathbf{I}_{ij}/\gamma, i, j = 1, \ldots, N$.
3) To calculate the matrices $\mathbf{A}$ and $\mathbf{B}$ using (10) and the $N \times N$ matrix $\mathbf{Q}$ using $\mathbf{A}(\mathbf{I} - \vec{1}\mathbf{B}^{\mathrm{T}})$.
4) To get the central row vector of matrix $\mathbf{Q}$ and reshape it into a weight kernel and then obtain the support value filter.

If the size of the mapped vector space is set to be 5×5 pixels, the spread parameter σ^2 in Gausian RBF kernel is set to be 0.3 and the γ of the LS-SVM is set to be 1 [19], the deduced support value filter is given as shown in (13), at the bottom of the following page.

E. Support Value Transform

We consider the image processing problem within LS-SVM framework. The support value image, support value of pixel defined in the mapped neighborhood center instead of conventional intensity value, is considered as an actual measure of the salient features of original image. Similar to conventional multiscale analysis method, a multiscale support value analysis

framework should be constructed for effective and complete extraction of these salient features.

1) Basic Multiscale Frameworks: The existing multiscale analysis tools can be divided into two basic schemes, pyramid and parallelepiped. The pyramid framework is widely used in the existing image fusion methods, such as the LAP [3], and the DWT methods [2], [5]–[7]. Another scheme is the parallelepiped framework, which is used in various undecimated wavelet transforms. An example is the "àtrous" algorithm [20]. Since the pyramid scheme includes a subsampling or decimating process, it may cause some distortions in the results, which appear obvious in the standard discrete wavelet transform. Therefore, the parallelepiped framework is considered as the support value analysis framework in this study. For this purpose, we first review a classical multiscale analysis method using parallelepiped framework, the undecimated wavelet transform—"àtrous" algorithm.

In the undecimated wavelet transform—"àtrous" algorithm [20]—the image decomposition scheme cannot be represented with a pyramid as in Mallat's algorithm (standard wavelet transform) but with a parallelepiped. The basis of the parallelepiped is the original image, A_2^j at a resolution 2^j. Each level of the parallelepiped is an approximation to the original image, as in Mallat's algorithm.

When climbing up through the resolution levels, the successive approximation images have a coarser spatial resolution but the same number of pixels as the original image. If a dyadic decomposition approach is applied, the resolution of the approximation image at the nth level is 2^{j-n}. These approximation images are computed using scaling functions. The spatial detail that is lost between the images A_2^{j-1} and A_2^j is collected in just one wavelet coefficient image, w_2^{j-1}, frequently called wavelet plane. This wavelet plane, which globally represents the horizontal, vertical and diagonal spatial detail between 2^j and 2^{j-1} resolution, is computed as the difference between A_2^{j-1} and A_2^j.

In contrast to Mallat's algorithm, all the approximation images obtained by applying this decomposition have the same number of columns and rows as the original image. This is a consequence of the fact that the "àtrous" algorithm is a nonorthogonal, redundant over-sampled transform. For the practical implementation of the "àtrous" algorithm, a 2-D filter associated to the scaling function is used. In order to obtain coarser approximations of the original image, the above filter must be filled with zeros for matching the resolution of desired level.

2) Support Value Transform-Parallelepiped Scheme: From the above analysis, we can see that the pyramid framework is to subsample or decimate the image while keep the filter constant, and the parallelepiped framework is to over-sample or fill the filter with zeros while keep the image size unchanged. As the parallelepiped framework is superior to the pyramid scheme in shift-invariance, it is suitable for the image fusion [20]. If we consider developing a support value analysis method based on the parallelepiped framework, it is necessary to introduce the spatial scale function into the support value computation procedure. However, the support value filter deduced from the mapped LS-SVM is a 2-D filter, and is indirectly associated with the scaling function. To introduce the spatial scale function into the deduced support value filter and generate a series of multiscale support value filters, introducing the spatial scale factor into the mapped LS-SVM and filling zeros in the basic support value filter to match the resolution of desired level are two potential techniques.

1) To introduce the spatial scale factor into the mapped LS-SVM.
   The support value filter deduced from the mapped LS-SVM is determined by the kernel function and the mapped input vector space. Therefore, two solutions respectively related to kernel function and input space may be feasible.
   - Solution based on scale-related input space. In the mapped LS-SVM, the mapped input vector space $\{(dr, dc):|dr| \leqslant m, |dc| \leqslant n\}$ does not include the scaling factor, which means that the interval between any adjacent input vectors is only one. To introduce the scale factor into the input space, we may define the interval between any adjacent input vectors is 2^{j-1}, where the j is the scale number, and the label vector $\mathbf{Y}$ correspondingly represents the label set of the input vector $\mathbf{x}$ with adjacent interval of 2^{j-1}. With this definition, the values of the kernel function $K(\mathbf{x}, \mathbf{x}_i)$ may indirectly include the scale factor. In this way, a multiscale mapped LS-SVM may be created and used to calculate the multiscale support values.
   - Solution based on scale-related kernel function. The kernel function of the LS-SVM may be scale-related. For example, Gaussian RBF kernel can be scale-related by setting the spread parameter σ [18], [19]. The mapped LS-SVM with RBF kernel having different values of spread parameter σ has different characteristics, and results in different support values.
2) To fill zeros in the basic support value filter.

As the first method analyzed above is indirect, we consider the second technique in this study. To match the resolution of desired level, the basic support value filter is filled with zeros. Based on this approach, a series of multiscale support value filters can be obtained. With the multiscale filters, a sequence of salient feature images can be extracted from the original image. Considering this solution is very simple and practically feasible for the support value analysis, we only discuss this approach in the following sections.

$$\begin{bmatrix} -0.0158 & -0.0136 & -0.0102 & -0.0136 & -0.0158 \\ -0.0136 & -0.0130 & -0.0602 & -0.0130 & -0.0136 \\ -0.0102 & -0.0602 & 0.5051 & -0.0602 & -0.0102 \\ -0.0136 & -0.0130 & -0.0602 & -0.0130 & -0.0136 \\ -0.0158 & -0.0136 & -0.0102 & -0.0136 & -0.0158 \end{bmatrix} \tag{13}$$

With the deduced basic support value filter, such as (13), a series of multiscale support value filters can be obtained by filling zeros in it, and then multiscale support value analysis is available, and the SVT-based image fusion method can be feasible.

III. Image Fusion Method Based on SVT

Similar to the "àtrous" wavelet transform, the SVT is a multiresolution transform with frame elements indexed by scale and location parameters. Unlike the "àtrous" wavelet transform, however, the support value transform directly provides the salient features of original image. It is an undecimated dyadic transform and is isotropic and shift-invariant and does not create artifacts. Given an image P, the sequence of its support values obtained by multiresolution decomposition is $\{S_1, S_2, \ldots, S_r\}$. The sequence of its approximations is the differences of the original image P and its support value image. That is

$$\begin{aligned} S_j &= SV_j * P_j \\ P_{j+1} &= P_j - S_j, \quad j = 1, \ldots, r, \quad P_1 = P. \end{aligned} \tag{14}$$

The reconstruction formula can be written as

$$P = P_{r+1} + \sum_{j=1}^{r} S_j \tag{15}$$

where r refers to the decomposition level, SV is the series of support value filters and $*$ represents convolution operation. To construct the sequence of support value images, the SVT algorithm performs successive convolutions with a filter SV obtained from the scaling function.

Similar to the traditional wavelet transform-based image fusion method, the general steps of the SVT-based image fusion approach can be listed as follows.

1) To co-register the original images and resample them to make their pixel size the same as one another.
2) To apply the SVT in (14) to each of the co-registered images and obtain a low-frequency component image $\{P_{r+1}\}$ and the support value image sequences $\{S_1, S_2, \ldots, S_r\}$.
3) To combine the multiple sets of low/support value components together. The most popular method is to select the components with the largest activity level at each pixel location [*choose-max* (CM)]. Alternatively, one can also take a weighted average of the different sets of components [*weighted average* (WA)], where the weights are determined by the activity level of the source components. Note that this approach does not require activity-level measurement and the support value components combination is a mere CM.
4) Optionally, to perform consistency verification, ensures that a fused component does not come from different source images from most of its neighbors. Usually, this is implemented by using a small majority filter [*window-based verification* (WBV)].
5) To use the inverse support value transform (ISVT) method in (15) to recover the fused image.

Fig. 1 illustrates the procedure of SVT-based image fusion on two original multifocus images, where the basic support value filter in (13) is used. First, the original images L and R are, respectively, decomposed into sequence of support value images plus the low-frequency components, as shown in L and R sequences in Fig. 1. Observing the decomposed image sequence, we can find that the SVT can effectively extract the salient features from the original image into the sequence of support value images. Second, the decomposed low-frequency components of L and R are merged into the resulting low-frequency components by averaging method, and the decomposed sequences of support value images are merged into the fused sequence of support value images by CM method, as shown in fused sequence in Fig. 1. Finally, the fused sequence of support value images and the low-frequency components are inversely transformed into the resulting image, as shown in Fig. 1. Carefully observing the original images L and R, we can see that the resulting image produced by the proposed fusion method is basically a combination of the in-focus parts of the source images. For example, both the word "edition" and character "Ope" appear clearly in the resulting image.

IV. Experimental Results

To verify the proposed image fusion method, we test it on the multisource images from the remote sensing image and various registered multisensor images from the collections of Manchester University and other organizations (provided by Dr. V. Petrovic),[1] which contains 120 group multisensor images. Six groups of the tested images are illustrated in Fig. 2. Rockinger's MATLAB toolbox[2] is used as the reference for the implementation of LAP method, standard DWT method and the undecimated "àtrous" wavelet (UAW) transform method. The window size of salience computation and the consistency check is set to be 3×3 pixels. For the SVT-based method, the basic support value filter in (13) is used.

All the experiments are implemented in Matlab 6.5 and run on a Pentium(R) 1.7-GHz machine with 512-M RAM.

A. Evaluation of Fusion Method

Evaluation of image fusion techniques is a nontrivial task especially as it is often difficult to say which of two fused images are better without reference specific tasks [21]. The existing evaluation techniques can be generally grouped into two categories: subjective or qualitative tests and objective or quantitative tests. Although the majority of evaluation of image fusion techniques tends to be subjective, this method is difficult to perform since it is based on psycho-visual testing and requires time and resources. Quantitative testing can overcome these difficulties, in theory, if metrics are validated. Although there have been many attempts to quantitatively test, as yet, no universally accepted standard has been developed for evaluating image fusion performance, the visual and quantitative analysis are used in this study.

To quantitatively evaluate the fusion performance, the following five assessment methods are considered in this study.

[1]http://www.ablen.com/hosting/imagefusion/resources/images/petrovic/petrovic.html

[2]http://www.stat.nctu.edu.tw/MISG/e_talk/medinfor/imagefushion/func.htm

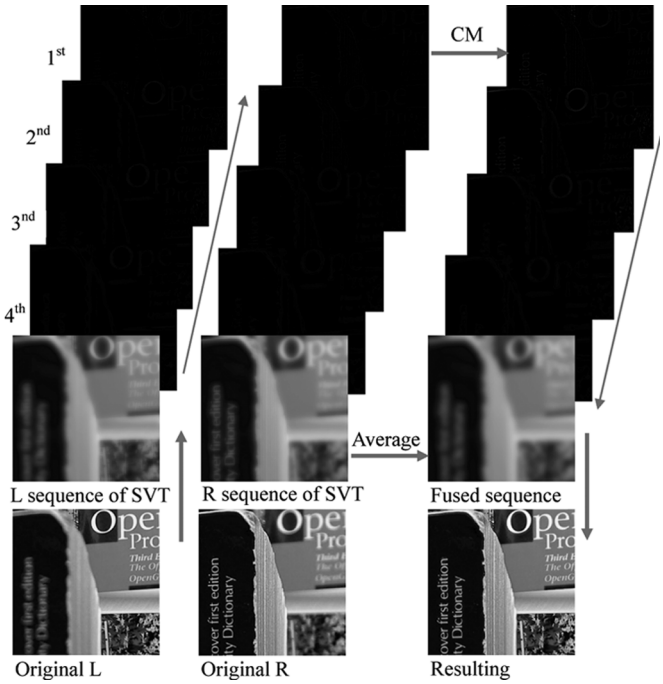


Fig. 1. Procedure of multifocus image fusion based on SVT.

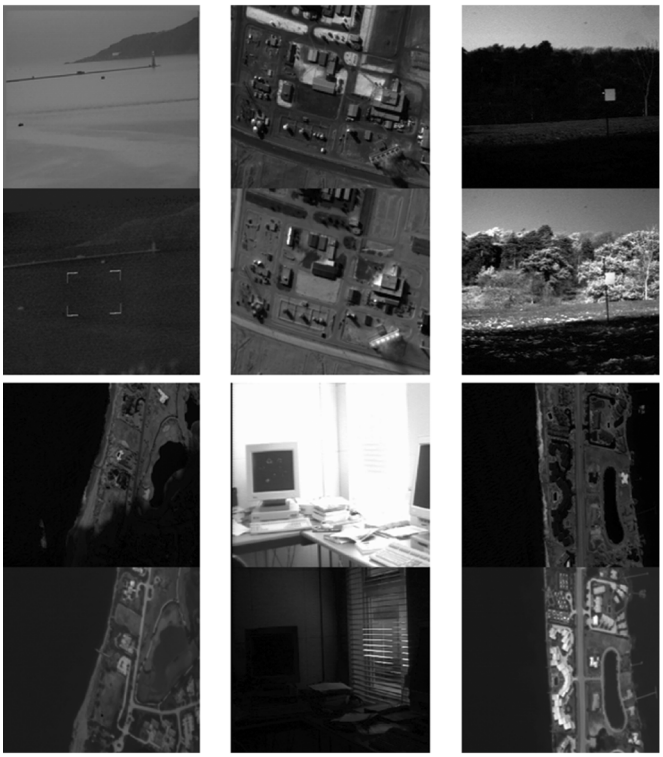


Fig. 2. Sample multisensor images for verifying the proposed fusion method.

1) Root mean squared error (RMSE) [22] of the difference image between the ideal image and the fused image, is an objective fidelity criterion to evaluate the multisource image fusion performance, and requires a ground truth fused image, which does not usually exist for real images. It should be close to 0 as much as possible.
2) Mutual information (MI) [23] essentially computes how much information from each of the original images is transferred to the fused image. The larger the MI value, the better the fused result is.
3) The Xydeas and Petrovic $Q^{AB/F}$ metric [24], considers the amount of edge information transferred from the input images to the fused images using a Sobel edge detector to calculate the strength and orientation information at each pixel in both source and the fused images. It should be close to 1 as much as possible.
4) Image quality metric (IQM) [25]-based metrics, are presented by Piella and Heijnmans [26], including the weighted fusion quality (WFQ) measure and edge dependent fusion quality index (EDFQ). The larger the WFQ and EDFQ values, the better the fused results are.
5) Condition entropy (CEntropy)[3] of fused results with original image is also used to estimate the information transferred from the input images. The larger the absolute CEntropy value, the better the fused result is.

B. Visual Inspection

Due to the large number of combinations that have been performed, and considering the research result [6], the LAP having advantageous performance over the conventional image fusion methods, we only visually inspect the fusion images obtained by using the LAP and SVT-based methods with setting of CM and averaging (AV), where the multiscale level is set to be fourth.

Fig. 3 illustrates the comparison of fused results on remote sensing images. Fig. 3(a) and (b), respectively, shows the original images pictured by different remote sensors, and Fig. 3(c) and (d) illustrates the resulting images, respectively, from the LAP and SVT-based image fusion approaches. Carefully observing Fig. 3(c) and (d), we can find that the proposed method provides more details than the LAP approach. For example, the road within the red circled region in Fig. 3(c) looks blurring, but it appears clear in Fig. 3(d).

The visual comparison of the fused results from the LAP and SVT-based methods on the sampled group images shown in Fig. 2 is illustrated in Fig. 4. For each group images, the upper fused image is obtained by using the LAP method and the lower result is obtained by using the proposed method. If we carefully observe these images, it can be found that the visual difference between the fused image from the LAP method and that from the SVT-based approach is not obvious.

From the above observation, we can see that the proposed image fusion method provides a little better than (as shown in Fig. 3) or a very close visual performance to the LAP method on multisensor images.

C. Quantitative Analysis

Table I gives the values of quality of visual information ($Q^{AB/F}$), mutual information (MI), image quality metric (IQM), IQM-based weighted fusion quality measure (WFQ), IQM-based edge dependent fusion quality index (EDFQ) and condition entropy (CEntropy) for the image fusion approaches including LAP, proposed SVT, UAW, and DWT image fusion

[3]http://www.cs.rug.nl/~rudy

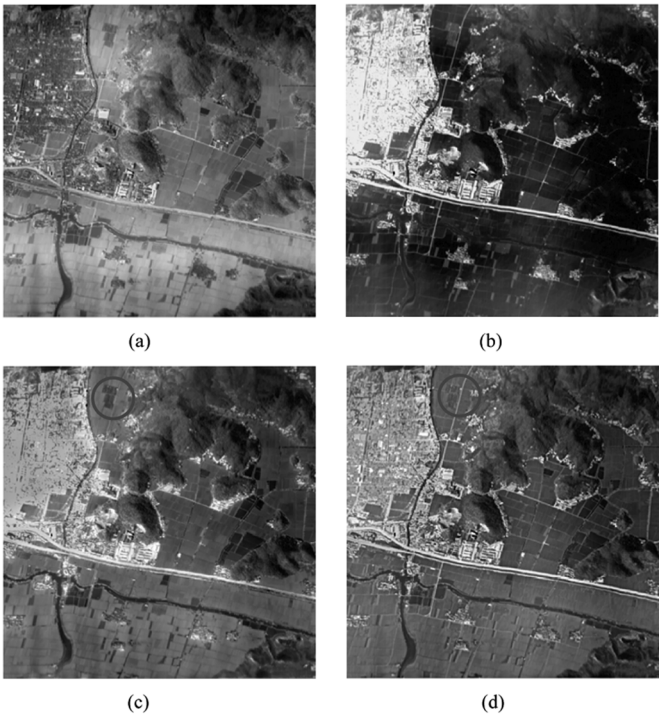


Fig. 3. Fused results from LAP method and SVT approach (low-frequencies: AV; salient features: CM).

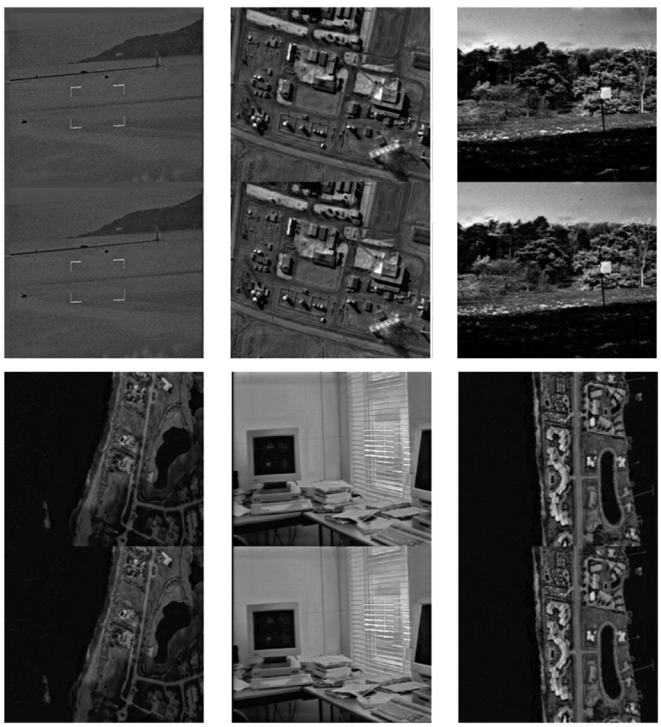


Fig. 4. Visual comparison of the fused results from the LAP and SVT-based methods on the sampled group images shown in Fig. 2. For each group image, the upper image is from the LAP method and the lower result is from the SVT-based method.

schemes, where the decomposition level is set to be fourth. All the reported values are the average results of the tested 120 group multisensor images.

TABLE I
COMPARISON OF FUSED RESULTS FROM DIFFERENT METHODS

	Features	$Q^{AB/F}$	MI	IQM	WFQ	EDFQ	CEntropy
a	LAP	0.7782	3.0101	0.9142	0.9499	0.9383	-4.5502
	SVT	0.7837	3.0437	0.9196	0.9530	0.9387	-4.5838
	UAW	0.7750	3.0020	0.9152	0.9502	0.9359	-4.5421
	DWT	0.7551	2.8683	0.9044	0.9453	0.9338	-4.4085
b	LAP	0.7907	3.0617	0.9168	0.9534	0.9397	-4.6018
	SVT	0.7926	3.0795	0.9213	0.9551	0.9407	-4.6196
	UAW	0.7844	3.0468	0.9171	0.9528	0.9370	-4.5869
	DWT	0.7649	2.9601	0.9063	0.9478	0.9337	-4.5002
c	LAP	0.7813	3.0260	0.9109	0.9494	0.9360	-4.5661
	SVT	0.7879	3.0458	0.9186	0.9532	0.9392	-4.5859
	UAW	0.7782	3.0063	0.9137	0.9504	0.9356	-4.5464
	DWT	0.7463	2.8663	0.8956	0.9425	0.9283	-4.4064

Low-frequencies: average (AV)
Coefficients selection for salient features: a.*Choose-max* (CM),
b. Salience match (*th.*=0.75), c. CM with consistency check

From Table I, we can first note that the optionally consistency verification performs, such as the salience match and the CM with consistency check, can effectively improve the fusion results in terms of the quantitative evaluation indexes, such as the $Q^{AB/F}$, MI, EDFQ and CEntropy, etc. For example, for the SVT-based method, the $Q^{AB/F}$ value is increased from 0.7837 to 0.7926, the MI value is increased from 3.0437 to 3.0795, and the EDFQ value is increased from 0.9387 to 0.9407, when the salient feature selection method changes from CM to salience match. Similar results can be found for other methods.

From Table I, we can note secondly that the proposed SVT-based method provides the best fusion performance in terms of resulting values of the quantitative evaluation indexes including $Q^{AB/F}$, MI, WFQ, EDFQ and CEntropy. For example, when the salient features selection method is CM, the $Q^{AB/F}$ value of the SVT method is 0.7837 and is greater than 0.7782, 0.7750, and 0.7551, respectively, provided by the LAP, UAW, and DWT methods. The MI value of the SVT method is 3.0437 and is greater than 3.0101, 3.0020, and 2.8683, respectively, provided by the LAP, UAW, and DWT approaches. The WFQ value of the SVT method is 0.9530 and is greater than 0.9499, 0.9502, and 0.9453, respectively, provided by the LAP, UAW, and DWT approaches. The EDFQ value of the SVT method is 0.9387 and is greater than 0.9383, 0.9359, and 0.9338, respectively, provided by the LAP, UAW, and DWT approaches. The absolute CEntropy value of the SVT method is 4.5838 and is greater than 4.5502, 4.5421, and 4.4085, respectively, provided by the LAP, UAW, and DWT approaches. The similar results can be found when the salient feature selection approach includes the optimally salience matching operation or the CM with consistency check operation. Based on the above analysis, we can see that the proposed SVT-based image fusion method is effective and is superior to the conventional image fusion methods including the LAP, UAW, and DWT methods, in terms of the related quantitative fusion evaluation indexes including $Q^{AB/F}$, MI, WFQ, EDFQ, and CEntropy, on the multisensor image fusion test.

D. Computation Efficiency Analysis

In order to estimate the computation efficiency of different image fusion methods, the time costs (seconds) for different approaches to fuse the 120 group multisensor images are listed in

TABLE II
COMPARISON OF COMPUTATION EFFICIENCY OF FUSION METHODS (SECONDS)

Feature selection	LAP	SVT	UAW	DWT
Choose-max (CM)	0.3614	0.5241	0.5974	0.6193
Salience match (*th*.=0.75)	0.6225	1.3479	1.4202	0.7944
CM with consistency check	0.6225	1.2936	1.3308	0.8170

Low-frequencies: average (AV)

Table II. All the values are the average results of the tested 120 group multisensor images.

From the data in Table II, we can see that the proposed SVT-based method is almost half as much again the time cost of the LAP method, when the salience features are selected by CM. The main reason is that the SVT is an undecimated transform, and needs more time to compute the support value images having the same size of the original image. Therefore, its computation efficiency is lower than the decimated methods including the LAP and DWT.

V. CONCLUSION

In the above analysis of the quantitative and visual inspection based on the remote sensing images and tested 120 group multisensor images released by Dr. V. Petrovic, we found that there are a little statistical differences between fused images from the proposed SVT-based image fusion method and the results from the LAP, DWT, and UAW methods in terms of the related fusion evaluation indexes. The proposed SVT-based method provides a little higher evaluation index values than the conventional image fusion methods including the LAP, UAW, and DWT methods although it has a little lower computation efficiency since it is an undecimated multiscale transform. Based on evaluation index value comparison of various image fusion methods, and combined the visual inspection, we can draw a conclusion that the support value can effectively represent the salient features of original images and the proposed SVT-based image fusion method can effectively integrate the salient features of the original images into the resulting image.

It is an alternative solution to effectively fuse multifocus and multisensor images. The effectively salient feature extraction is a developing problem in image processing such as the image fusion, and we made an exploration in representation and extraction of the image's salient features. We introduced the support value into the image fusion, and developed a new multiscale analysis tool, and further extend the application areas of the support vector machines.

REFERENCES

[1] S. T. Li, J. T. Y. Kwok, I. W. H. Tsang, and Y. N. Wang, "Fusing images with different focuses using support vector machines," *IEEE Trans. Neural Netw.*, vol. 15, no. 6, pp. 1555–1561, Jun. 2004.

[2] P. K. Varshney, "Multisensor data fusion," *Electron. Commun. Eng. J.*, vol. 9, no. 12, pp. 245–253, 1997.

[3] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. Ogden, "Pyramid methods in image processing," *RCA Eng.*, vol. 29, no. 6, pp. 33–41, 1984.

[4] S. Li, J. T. Kwok, and Y. Wang, "Multifocus image fusion using artificial neural networks," *Pattern Recognit. Lett.*, vol. 23, no. 8, pp. 985–997, 2002.

[5] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.

[6] G. Pajares and J. M. Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, 2004.

[7] V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 228–237, Feb. 2004.

[8] L. Bogoni and M. Hansen, "Pattern-selective color image fusion," *Pattern Recognit.*, vol. 34, no. 8, pp. 1515–1526, 2001.

[9] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.

[10] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[11] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.

[12] S. Zheng, Y. Q. Sun, J. W. Tian, and J. Liu, "Mapped least squares support vector regression," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 19, no. 3, pp. 459–475, 2005.

[13] S. Zheng, J. Liu, and J. W. Tian, "A new efficient SVM-based edge detection method," *Pattern Recognit. Lett.*, vol. 25, no. 10, pp. 1143–1154, 2004.

[14] S. Zheng, J. W. Tian, J. Liu, and C. Y. Xiong, "Novel algorithm for image interpolation," *Opt. Eng.*, vol. 43, no. 4, pp. 856–865, 2004.

[15] S. Zheng, J. Liu, and J. W. Tian, "An efficient star acquisition method based on SVM with mixtures of kernels," *Pattern Recognit. Lett.*, vol. 26, no. 2, pp. 147–165, 2005.

[16] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: Robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1, pp. 85–105, 1999.

[17] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations," *Philos. Trans. Roy. Soc. Lond. A*, no. 209, pp. 415–446, 1909.

[18] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, vol. 2, no. 2, pp. 321–355, 1988.

[19] S. Zheng, J. Liu, and J. W. Tian, "Research of mapped least squares SVM optimal configuration," in *Applied Soft Computing Technologies: The Challenge of Complexity, Vol. Advances in Soft Computing*, A. Abraham, B. D. Baets, M. Köppen, and B. Nickolay, Eds. Berlin, Germany: Springer, 2006, pp. 685–694.

[20] M. Gonzalez-Audicana, X. Otazu, O. Fors, and A. Seco, "Comparison between Mallat's and the "à trous" transform based algorithms for the fusion of panchromatic images," *Int. J. Remote Sens.*, vol. 26, no. 3, pp. 595–614, 2005.

[21] J. L. John, J. O. Robert, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Inf. Fusion*, vol. 8, no. 2, pp. 119–130, 2007.

[22] H. Li, S. Manjunath, and S. Mitra, "Multi sensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, 3, pp. 235–245, 1995.

[23] G. H. Qu, D. L. Zhang, and P. F. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, 2002.

[24] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, Feb. 2000.

[25] Z. Wang and A. Bovik, "A universal image quality index," *Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, 2002.

[26] G. Piella and H. Heijmans, "A new quality metric for image fusion," presented at the Int. Conf. Image Processing, Barcelona, Spain, 2003.

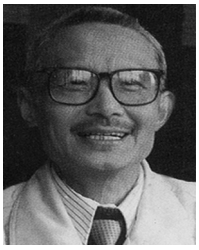
Sheng Zheng received the M.S. degree in wireless electronics from Huazhong Normal University, China, in 1992, and the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2005.

He was a Visiting Scholar at The Hong Kong Polytechnic University, Hong Kong, from January to May 2006. He is currently a Postdoctorate in communication and information system at HUST and a Professor and M.S. Supervisor in computer science and its applications in the College of Electrical and Information Technology, and the Vice Director of the Institute of Intelligent Vision and Image Information Best, China Three Gorges University. His research interests include image processing, object recognition, computer vision, neural networks, and intelligent control. He has published more than 40 papers.

Wen-Zhong Shi received the Ph.D. degree from the University Osnabrueck, Vechta, Germany, in 1994.

He is a Professor in GIS and remote sensing with the Department of Land Surveying and Geo-Informatics and Director of the Advanced Research Centre for Spatial Information Technology, The Hong Kong Polytechnic University, Hong Kong. He has been researching and working in the area of GIS and remote sensing since 1985 in Hong Kong, Germany, the Netherlands, Australia, and mainland China. He has published over 300 research articles (over 50 are in SCI journals) and ten books. His current research interests include GIS, remote sensing, uncertainty of spatial data and quality control, image processing for high resolution satellite images, 3-D and dynamic data model in GIS, virtual reality, design and development of GIS, and the integration of GIS and remote sensing.

Dr. Shi is the Chair of WG II-7, the International Society for Photogrammetry and Remote Sensing, and he was the President for the Hong Kong Geographic Information System Association (2001–2003).



Jian Liu received the degree from the Radio Technology Department, Huazhong University of Science and Technology (HUST), Wuhan, China, in 1961.

He was a Visiting Scholar at the Ecole Nationale Superieure des Telcommunications (ENST), France, from 1979 to 1981, where he was also a Guest Professor from 1989 to 1990. He is currently a Professor and Ph.D. Supervisor of pattern recognition and artificial intelligence at HUST. He has published more than 150 papers. He is the co-compiler of two books: *Digital Image Processing* and *Remote Sensing Image Digital Processing*, which was awarded the first prize of Excellent Textbooks by the Ministry of Mechanical and Electronics Industries. His main research fields include digital image processing and recognition, computer vision, and remote sensing image analysis.

Dr. Liu is a member of the Remote Sensing, Telemetry, and Remote Control Committee of the Chinese Radio Engineering Society, and a senior member of the Chinese Electronics Society. He has accomplished 20 research projects, among which six were awarded prizes for Scientific and Technological Progress at the provincial or ministerial level.

Guang-Xi Zhu received the degree in radio engineering from the Huazhong Institute of Technology, China, in 1969.

He is a Professor and Ph.D. Supervisor of Electronics and Information Engineering, Chinese Key Laboratory of Optic and Electronics, Huazhong University of Science and Technology (HUST), Wuhan, China. He has published more than 150 papers. His main research fields include broadband mobile communication, image processing, etc.

Jin-Wen Tian received the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 1998.

He is a Professor and Ph.D. Supervisor for pattern recognition and artificial intelligence, State Education Commission Key Laboratory for Image Processing and Intelligent Control, HUST. His main research topics are remote sensing image analysis, wavelet analysis, image compression, computer vision, and fractal geometry. He has published more than 60 papers.