# A MAP Approach for Joint Motion Estimation, Segmentation, and Super Resolution

Huanfeng Shen, Liangpei Zhang, Bo Huang, and Pingxiang Li

***Abstract*—Super resolution image reconstruction allows the recovery of a high-resolution (HR) image from several low-resolution images that are noisy, blurred, and down sampled. In this paper, we present a joint formulation for a complex super-resolution problem in which the scenes contain multiple independently moving objects. This formulation is built upon the maximum *a posteriori* (MAP) framework, which judiciously combines motion estimation, segmentation, and super resolution together. A cyclic coordinate descent optimization procedure is used to solve the MAP formulation, in which the motion fields, segmentation fields, and HR images are found in an alternate manner given the two others, respectively. Specifically, the gradient-based methods are employed to solve the HR image and motion fields, and an iterated conditional mode optimization method to obtain the segmentation fields. The proposed algorithm has been tested using a synthetic image sequence, the "Mobile and Calendar" sequence, and the original "Motorcycle and Car" sequence. The experiment results and error analyses verify the efficacy of this algorithm.**

***Index Terms*—Joint estimation, maximum *a posteriori* (MAP), motion estimation, segmentation, super resolution.**

## I. Introduction

HIGH-RESOLUTION (HR) images have a variety of applications in remote sensing, video frame freezing, medical diagnostics, and military information acquisition. However, low-resolution (LR) images are used more frequently due to the high cost and physical constraints of the hardware. Consequently, super-resolution (SR) reconstruction has emerged as an alternative for producing one or a set of HR images from a sequence of LR images.

The multiframe SR problem was first formulated by Tsay and Huang [1] in the frequency domain. Their formulation was extended by Kim *et al.* [2], [3] to consider observation noise as well as spatial blurring. A discrete cosine transform (DCT) based frequency domain method was developed in [4]. In general, the frequency domain approaches have the strength of theoretical simplicity and high computational efficiency. Such methods are limited in their approach, as they are not able to accommodate spatial domain *a priori* knowledge, nonglobal translational motion models, or spatially varying degradation [5]. Consequently, many kinds of spatial domain approaches have been developed to overcome the weaknesses of the frequency domain approaches. Typical spatial domain approaches include nonuniform interpolation [6] based on the generalized multichannel sampling theorem [7], [8], iterative back projection (IBP) [9], projection onto convex sets (POCS) [10]–[13], maximum likelihood (ML) [14], maximum a posteriori (MAP) [15], [16], hybrid ML /MAP/POCS [17], and adaptive filtering [18].

The fundamental notion of SR is to employ either frequency domain or spatial domain approaches to bring together the additional information in each image provided by the subpixel displacements from a referenced image. Before reconstructing one frame, this frame must be fixed as a referenced frame and then all other frames are registered against it (It is noted that in this paper, the motion field describes the motions between a current frame and the referenced frame). Hence, motion estimation/registration plays a critical role in SR image reconstruction. Work has been carried out with emphasis on reducing the effect of registration error in SR. Bose *et al.* [19] suggested a total least squares method to improve the solution accuracy when errors exist in both observations as well as the system matrix. Ng and Bose [20] gave an analysis of the motion errors on the convergence rate of the iterative approach for solving the transform based preconditioned system of equations. Lee and Kang [21] proposed an approach based on channel adaptive regularization to minimize the effect of the registration error. Shah and Zakhor [22] considered the registration inaccuracy by finding a set of candidate motion estimates for each pixel. As an alternative solution, the simultaneous registration and reconstruction approach reduces the effect of registration error. Tom and Katsaggelos [23] developed a simultaneous registration and reconstruction approach, in which they formulated the SR problem in the form of a Maximum Likelihood problem and solved it using the expectation–maximization algorithm. Russell *et al.* [24] developed an approach within the MAP framework to simultaneously estimate the image registration parameters and the HR image. Segall *et al.* [25], [26] presented a noteworthy approach the involved joint estimation of dense motion vectors and HR image for compressed video. More recently, Woods *et al.* [27] presented the complex stochastic methods in which the parameters of registration, noise and image statistics are jointly estimated based on the available observations. Using the Gauss–Newton method, Chung *et al.* [28] attempted completely coupled and partially coupled optimizations to improve the efficiency of their solution.

Nevertheless, for the most part, SR methods [19]–[28] have not adequately addressed the more complex multiple moving

Manuscript received March 29, 2006; revised August 5, 2006. This work was supported in part by the National Natural Science Foundation of China under Grant 40471088, 40523005, in part by the 973 Program of China under Grant 2006CB701302, and in part by the Foundation of State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing under Grant 904151695. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Michael Elad.

H. Shen, L. Zhang, and P. Li are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@public.wh.hb.cn).

B. Huang is with the Geography and Resource Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong.

Digital Object Identifier 10.1109/TIP.2006.888334

objects problem. When scenes contain multiple independently moving objects, the estimated motion vectors are prone to be inaccurate around the motion boundaries and occlusion regions, which results in artifacts in the reconstructed HR image. SR enhancement of arbitrary scenes is the ultimate goal of such research effort [29]. Thus, Irani and Peleg [30] proposed an object-based approach in which the multiple moving objects are first detected using a tracking algorithm recursively. Subsequently, they employed a back-projection algorithm to improve the resolution of different objects. This approach was extended by Eren *et al.* [31] based on the POCS framework. They assumed that objects of interest are marked on a reference LR frame interactively. These are then tracked on all available frames using a mesh-based object-tracking method. Finally, a POCS approach was used to realize the SR. For the SR reconstruction of compressed video sequence, Alvarez *et al.* [32] attempted a method to determine which pixels provide useful information to estimation of the HR image, and considered only those observations by modifying the acquisition model.

In this paper, we propose a novel joint estimation approach for motion estimation, segmentation, and SR reconstruction to deal with the multiple moving objects problem. Our approach is based on the following recognitions:

The desired HR image and motion estimates are interdependent. Accurate subpixel motion estimates are critical for SR image reconstruction. On the other hand, a high-quality HR image can also facilitate accurate motion estimates.

Motion estimation and segmentation are also interdependent. The success of motion segmentation is closely related to the accuracy of the motion field, and vice versa [33].

Motion segmentation can benefit the SR result, especially for scenes containing multiple independent moving objects.

As motion estimation, segmentation and SR reconstruction are mutually interdependent and influence each other, an ideal approach is addressing them simultaneously. To the best of our knowledge, only two among these three processes have been simultaneously addressed, such as the joint estimation of motion and reconstruction [23]–[28] and that of motion and segmentation [33], [34]. In our algorithm, all the three processes are judiciously integrated within a MAP framework. The motion fields and segmentation fields are iteratively updated along with the HR image in a cyclic optimization procedure. The algorithm reinforces the interdependence among the motion estimates, segmentation map and HR image in a mutually beneficial manner. In particular, it can suppress the artifacts around motion boundaries and occlusion regions without the need of interaction.

The remainder of the paper is organized as follows. In Section II, the SR observation model and motion field model are described. The joint MAP estimation problem is formulated in Section III. In Section IV, the joint optimization procedure to solve the motion fields, segmentation fields and HR image is presented. Experimental results are provided in Section V, and Section VI concludes the paper.

## II. Models

Motion estimation, segmentation, and SR involve the basic image observation model and the motion field model. The image observation model relates the original HR image to the observed LR images. A simple but efficient observation model is crucial for analyzing the SR problem in a comprehensive manner. The motion field model is a description or representation of a motion field, which is critical for motion estimation and segmentation. This section introduces a general image observation model at first and then describes a motion field model.

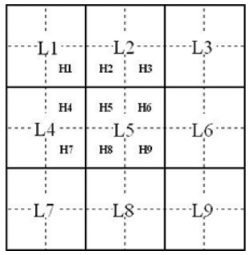


Fig. 1. Down-sampling grid.

### A. Image Observation Model

Typically, the imaging process involves warping followed by blurring and down sampling to generate LR images from an HR image. Let the underlying HR image be denoted in the vector form by $\boldsymbol{z} = [z_1, z_2 \ldots, z_{L_1N_1 \times L_2N_2}]^T$, where $L_1N_1 \times L_2N_2$ is the HR image size. Letting $L_1$ and $L_2$ denote the down-sampling factors in the horizontal and vertical directions, respectively, each observed LR image having the size $N_1 \times N_2$. Thus, the LR image can be represented as $\boldsymbol{y}_k = [y_{k,1}, y_{k,2} \ldots, y_{k,N_1 \times N_2}]^T$, where $k = 1, 2, \ldots P$, with $P$ being the number of the LR images. Assuming that each observed image is contaminated by additive noise, the observation model can be represented as [5], [17], [35]

$$\boldsymbol{y}_k = \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z} + \boldsymbol{n}_k \tag{1}$$

where $\boldsymbol{M}_k$ is the warp matrix with the size of $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, $\boldsymbol{B}_k$ represents the blur matrix also of size $L_1N_1L_2N_2 \times L_1N_1L_2N_2$, $\boldsymbol{D}$ is a $N_1N_2 \times L_1N_1L_2N_2$ down-sampling matrix, and $\boldsymbol{n}_k$ represents the $N_1N_2 \times 1$ noise vector. It is assumed in this paper that the noise is a zero mean Gaussian type, and the optical blur functions are uniform and to be known. The down-sampling is implemented using an average strategy, with which the LR pixel L5 is equal to the average of the four HR pixels H5, H6, H8, and H9 in the down-sampling grid shown in Fig. 1.

### B. Motion Field Model

The following motion field model is drawn from [33], [34]. Let us denote a 2-D motion field as $\boldsymbol{m}_k$ and a motion vector as $\boldsymbol{m}_k(\boldsymbol{x}) = [m_{k,u}(\boldsymbol{x}), m_{k,v}(\boldsymbol{x})]$, where $\boldsymbol{x} = [x_u, x_v]$ is the pixel site. If the scene consists of $R$ independently moving objects, then there should exist an underlying segmentation label $\boldsymbol{l}_k$ which assigns each motion vector to one of the $R$ classes. Generally, each object can be approximated by a parametric model, such as affine transformation and perspective model. Thus, the motion field $\boldsymbol{m}_k$ can be represented as the sum of a parametric motion field $\tilde{\boldsymbol{m}}_k$ and a nonparametric residual field $\boldsymbol{m}_k^{\boldsymbol{r}}$, which accounts for local motion and other modeling errors [36]; that is

$$\boldsymbol{m}_k(\boldsymbol{x}) = \tilde{\boldsymbol{m}}_k(\boldsymbol{x}) + \boldsymbol{m}_k^r(\boldsymbol{x}). \tag{2}$$

The following eight-parameter perspective models are selected:

$$\tilde{m}_{k,u}(\boldsymbol{x}) = a_0 + a_1 x_u + a_2 x_v + a_6 x_u^2 + a_7 x_u x_v$$
$$\tilde{m}_{k,v}(\boldsymbol{x}) = a_3 + a_4 x_u + a_5 x_v + a_6 x_u x_v + a_7 x_v^2 \quad (3)$$

where $a_0, a_1, \ldots a_7$ are the mapping parameters.

## III. Problem Formulation

Let the full set of $P$ LR images, motion fields, and segmentation fields be denoted by $\boldsymbol{y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots \boldsymbol{y}_P\}$, $\boldsymbol{m} = \{\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots \boldsymbol{m}_P\}$ and $\boldsymbol{l} = \{\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots \boldsymbol{l}_P\}$, respectively. The purpose is to realize the joint MAP estimate of motion fields $\boldsymbol{m}$, segmentation fields $\mathrm{l}$ and HR image $\mathbf{z}$, given the observed LR images $\boldsymbol{y}$. The estimate can be computed by

$$\hat{\boldsymbol{z}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \{p(\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l} \,|\, \boldsymbol{y})\}. \quad (4)$$

Applying Bayes' rule yields, (4) becomes

$$\hat{\boldsymbol{z}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \left\{ \frac{p(\boldsymbol{y} \,|\, \boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l}) p(\boldsymbol{z}, m, \boldsymbol{l})}{p(\boldsymbol{y})} \right\}. \quad (5)$$

Since $p(\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l} \,|\, \boldsymbol{y})$ is independent of $\boldsymbol{y}$, (5) can be rewritten as

$$\begin{aligned} \hat{\boldsymbol{z}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{l}} &= \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \{p(\boldsymbol{y} \,|\, \boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l}) p(\boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l})\} \\ &= \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \{p(\boldsymbol{y} \,|\, \boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l}) p(\boldsymbol{z}, \boldsymbol{m} \,|\, \boldsymbol{l}) p(\boldsymbol{l})\} \\ &= \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \{p(\boldsymbol{y} \,|\, \boldsymbol{z}, \boldsymbol{m}, \boldsymbol{l}) p(\boldsymbol{z} \,|\, \boldsymbol{l}, \boldsymbol{m}) p(\boldsymbol{m} \,|\, \boldsymbol{l}) p(\boldsymbol{l})\}. \end{aligned} \quad (6)$$

Assuming different $\boldsymbol{y}_k$, $\boldsymbol{m}_k$, and $\boldsymbol{l}_k$ are independent, respectively, we obtain

$$\begin{aligned} &\hat{\boldsymbol{z}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{l}} \\ &= \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \left\{ \prod_k [p(\boldsymbol{y}_k | \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k) p(\boldsymbol{z} \,|\, \boldsymbol{l}_k, \boldsymbol{m}_k) p(\boldsymbol{m}_k | \boldsymbol{l}_k) p(\boldsymbol{l}_k)] \right\}. \end{aligned} \quad (7)$$

Using the monotonic log function, it can be expressed as (8), shown at the bottom of the page.

At the outset, we define two diagonal matrices which will be used to set up some of the four PDF functions in (8). The presence of occlusion regions and the restriction of motion estimation methods lead to pixels that cannot be predicted from the referenced HR image and that have poor and conflicting motion estimates in the un-referenced HR images. The same term "unobservable" is used to denote the two types of pixels as that in [32]. To denote the observable pixels in the HR image $\boldsymbol{z}_k = \boldsymbol{M}_k \boldsymbol{z}$, a diagonal matrix $\boldsymbol{O}_k^h$ with the size of $L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2$ is defined. Each HR pixel in $\boldsymbol{z}_k$ corresponds to a diagonal element in $\boldsymbol{O}_k^h$. If a HR pixel is observable, its corresponding diagonal element is 1, otherwise 0. Once a pixel in HR image $\boldsymbol{z}_k$ is detected as unobservable, all the pixels in the corresponding LR image $\boldsymbol{y}_k$ dependent on it using the observation model (1) are also considered unobservable. Another $N_1 N_2 \times N_1 N_2$ diagonal matrix $\boldsymbol{O}_k^l$ is defined with the similar signification as $\boldsymbol{O}_k^h$. The method to detect the pixels that are observable or unobservable will be presented in Section IV (C).

The first probability density function (PDF), $p(\boldsymbol{y}_k \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k)$, provides a measure of the conformance of the present HR image, motion field and segmentation field to the observed LR frame $k$. $\boldsymbol{y}_k'$ and $\boldsymbol{y}_k''$ denoting the vectors containing the observable and unobservable pixels, respectively, the PDF can be calculated as follows:

$$p(\boldsymbol{y}_k \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k) = p(\boldsymbol{y}_k' \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k) p(\boldsymbol{y}_k'' \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k). \quad (9)$$

Since the unobservable pixels should not be considered in SR reconstruction, $p(\boldsymbol{y}_k'' \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k)$ can be regarded as a constant. The central task then comes down to defining $p(\boldsymbol{y}_k' \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k)$. One choice is to determine $\boldsymbol{y}_k'$ and its corresponding down-sampling matrix $\boldsymbol{D}'$ first, then $\boldsymbol{y}_k'$ and $\boldsymbol{D}'$ are used to give the PDF function [15], [32]. An alternative method is that the original form of $\boldsymbol{y}_k$ and $\boldsymbol{D}$ are used by affiliating the diagonal matrix $\boldsymbol{O}_k^l$ into the PDF function, given by

$$p(\boldsymbol{y}_k \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k) = \frac{1}{C_1} \exp\left\{ -\frac{\left\| \boldsymbol{O}_k^l (\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k \boldsymbol{M}_k \boldsymbol{z}) \right\|^2}{2\sigma_k^2} \right\} \quad (10)$$

where $C_1$ is a constant, and $\sigma_k^2$ is the error variance.

The second density function $p(\boldsymbol{z} \,|\, \boldsymbol{l}_k, \boldsymbol{m}_k)$ models the image prior. In most cases which consider the joint estimation of motion and HR image, it is assumed that this prior is independent [24]–[26]. In theory, it is possible to obtain some prior information from the segmentation field. For example, regions located at the segmentation boundaries, to some extent, tend to be sharper than other regions. However, it is hard to utilize such information for improving the SR result because there is a high requirement on the segmentation accuracy. Hence, it is assumed the HR image prior is independent of motion field $\boldsymbol{m}_k$ and segmentation field $\boldsymbol{l}_k$. In this paper, the following prior model is used:

$$p(\boldsymbol{z} \,|\, \boldsymbol{l}_k, \boldsymbol{m}_k) = p(\boldsymbol{z}) = \frac{1}{C_2} \exp\left\{ -\lambda_1 \left\| \boldsymbol{Q}_1 \boldsymbol{z} \right\|^2 \right\} \quad (11)$$

where $\lambda_1$ is a parameter that controls the variance of the prior distribution, and $\boldsymbol{Q}_1$ represents a linear high-pass operation that penalizes the estimation which is not smooth. $\boldsymbol{Q}_1$ is chosen as a 2-D Laplacian in this paper.

The *a priori* density of the motion $p(\boldsymbol{m}_k \,|\, \boldsymbol{l}_k)$ is modeled in such a way that there is an interaction between the motion field and the segmentation field

$$\begin{aligned} &p(\boldsymbol{m}_k \,|\, \boldsymbol{l}_k) \\ &= \frac{1}{C_3} \exp\left\{ -\lambda_2 \sum_{\boldsymbol{x}} \left\| \boldsymbol{O}_k^h (\boldsymbol{m}_k(\mathrm{x}) - \tilde{\boldsymbol{m}}_k(\mathrm{x})) \right\|^2 - \lambda_3 \left\| \boldsymbol{Q}_2 \boldsymbol{m}_k \right\|^2 \right\}. \end{aligned} \quad (12)$$

$$\hat{\boldsymbol{z}}, \hat{\boldsymbol{m}}, \hat{\boldsymbol{l}} = \arg\max_{\boldsymbol{z},\boldsymbol{m},\boldsymbol{l}} \left\{ \prod_k [\log p(\boldsymbol{y}_k \,|\, \boldsymbol{z}, \boldsymbol{m}_k, \boldsymbol{l}_k) + \log p(\boldsymbol{z} \,|\, \boldsymbol{l}_k, \boldsymbol{m}_k) + \log p(\boldsymbol{m}_k \,|\, \boldsymbol{l}_k) + \log p(\boldsymbol{l}_k)] \right\} \quad (8)$$

The first term in (12) aims to minimize the deviation of the motion field $\boldsymbol{m}_k(\mathrm{x})$ from the parametric motion field $\tilde{\boldsymbol{m}}_k(\mathrm{x})$; thus, it enforces a minimum norm estimate of the residual motion field $\boldsymbol{m}_k^r(\boldsymbol{x})$ [33]. It should be noted that the constraint is enforced only on the observable pixels by affiliating the diagonal matrix $\boldsymbol{O}_k^h$. The second term imposes spatial smoothness on the estimated motion vectors. The operator $\boldsymbol{Q}_2$ is an improved Laplacian, using which the spatial smoothness is not enforced on motion vectors generated by different objects in the neighborhood. The parameters $\lambda_2$ and $\lambda_3$ control the relative contribution of the two terms.

The last distribution function in (7) provides a prior probability of the segmentation field, which is modeled such that the object discontinuities coincide with spatial intensity boundaries [37]

$$p(\boldsymbol{l}_k) = \frac{1}{C_4} \exp\left\{ -\sum_{\boldsymbol{x_i}} \sum_{\boldsymbol{x_j} \in \wp} \boldsymbol{V}_C(l_k(\boldsymbol{x}_i), l_k(\boldsymbol{x}_j)) \right\} \quad (13)$$

where $\wp$ denotes the neighborhood system of $\mathbf{x}_i$, and the clique potential $\boldsymbol{V}_C(\cdot)$ encourages the stochastically estimated labels to match a predetermined spatial segmentation field

$$\boldsymbol{V}_C(l_k(\boldsymbol{x}_i), l_k(\boldsymbol{x}_j)) = \begin{cases} -\gamma, & \text{if } l_k(\boldsymbol{x}_i) = l_k(\boldsymbol{x}_j) \\ +\gamma, & \text{otherwise} \end{cases}. \quad (14)$$

Substituting (10)–(13) into (8), the minimization cost function is obtained in (15), shown at the bottom of the page.

This cost function is used in the optimization procedure introduced subsequently.

## IV. Optimization Procedure

As the cost function (15) consists of three large sets of unknowns (motion estimates, segmentation labels, and HR pixel values), using direct search techniques is intractable. Therefore, a cyclic coordinate descent optimization procedure is developed to solve the unknowns. The motion fields, label fields and HR image are found in an alternate manner given the two others, respectively.

### A. Updating the HR Image

Given a fixed motion field $\boldsymbol{m}_k$ and segmentation field $\boldsymbol{l}_k$, the parametric motion field $\tilde{\boldsymbol{m}}_k$ and the two diagonal matrices $\boldsymbol{O}_k^h$ and $\boldsymbol{O}_k^l$ can also be regarded as known quantities because they can be easily obtained (see Section IV-C). Thus, if the full set of motion fields $\boldsymbol{m}$ and segmentation fields $\boldsymbol{l}$ is known, the desired HR image can be updated by minimizing the following cost function

$$E_1(\boldsymbol{z}) = \sum_k \left\| \boldsymbol{O}_k^l(\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z}) \right\|^2 + \lambda_1 \left\| \boldsymbol{Q}_1\boldsymbol{z} \right\|^2 \quad (16)$$

which is composed of two terms in (15) that contain the sole unknown quantity $\boldsymbol{z}$. A gradient descent procedure is designed to minimize the cost function. Differentiating (16) with respect to $\mathbf{z}$ and setting the result equal to zero, we have

$$\nabla E_1(\boldsymbol{z}) = -2\sum_k \boldsymbol{M}_k^T\boldsymbol{B}_k^T\boldsymbol{D}^T(\boldsymbol{O}_k^l)^T\boldsymbol{O}_k^l(\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z}) + 2\lambda_1\boldsymbol{Q}_1^T\boldsymbol{Q}_1\boldsymbol{z} = 0. \quad (17)$$

Since $\boldsymbol{O}_k^l$ is diagonal, $(\boldsymbol{O}_k^l)^T\boldsymbol{O}_k^l = \boldsymbol{O}_k^l$. Thus, the HR image is solved by employing the successive approximations iteration

$$\hat{\boldsymbol{z}}^{n+1} = \hat{\boldsymbol{z}}^n + \alpha^n\boldsymbol{r}^n \quad (18)$$

where

$$\boldsymbol{r}^n = \sum_k \boldsymbol{M}_k^T\boldsymbol{B}_k^T\boldsymbol{D}_k^T\boldsymbol{O}_k^l(y_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\hat{\boldsymbol{z}}^n) - \lambda_1\boldsymbol{Q}_\mathbf{1}^T\boldsymbol{Q}_\mathbf{1}\hat{\boldsymbol{z}}^n \quad (19)$$

and $\alpha^n$ represents the step size at the $n$th iteration which is critical for the convergence. If it is too small, the convergence will be very slow. On the other hand, if it is too large, the algorithm will be unstable or divergent. Differentiating $E_1(\hat{\boldsymbol{z}}^n + \alpha^n\boldsymbol{r}^n)$ with respect to $\alpha^n$ and set the result equal to zero, after some manipulation, the optimal step size is solved by

$$\alpha^n = \frac{(\boldsymbol{r}^n)^T\boldsymbol{r}^n}{\sum_k \left\| \boldsymbol{O}_k^l\boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{r}^n \right\|^2 + \lambda_1 \left\| \boldsymbol{Q}_\mathbf{1}\boldsymbol{r}^n \right\|^2}. \quad (20)$$

### B. Updating the Motion Field

Given the estimate of the HR image $\boldsymbol{z}$ and segment field $\boldsymbol{l}_k$, the cost function to update the motion filed $\boldsymbol{m}_k$ can be given

$$E_2(\boldsymbol{m}_k) = \left\| \boldsymbol{O}_k^l(\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z}) \right\|^2 + \lambda_2 \sum_{\boldsymbol{x}} \left\| \boldsymbol{O}_k^h(\boldsymbol{m}_k(\boldsymbol{x}) - \tilde{\boldsymbol{m}}_k(\boldsymbol{x})) \right\|^2 + \lambda_3 \left\| \boldsymbol{Q}_2\boldsymbol{m}_k \right\|^2. \quad (21)$$

Since all the three terms are differentiable, the gradient-based optimization method is used again. Differentiating (21) with respect to $\boldsymbol{m}_k$ and setting the result equal to zero

$$\nabla \mathrm{E}_2(\boldsymbol{m}_k) = -2\frac{\partial \boldsymbol{M}_k\boldsymbol{z}}{\partial \boldsymbol{m}_k}\boldsymbol{B}_k^T\boldsymbol{D}^T(\boldsymbol{O}_k^l)^T\boldsymbol{O}_k^l(\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z}) + 2\lambda_2(\boldsymbol{O}_k^h)^T\boldsymbol{O}_k^h(\boldsymbol{m}_k(\boldsymbol{x}) - \tilde{\boldsymbol{m}}_k(\boldsymbol{x})) + 2\lambda_3\boldsymbol{Q}_2^T\boldsymbol{Q}_2\boldsymbol{m}_k = 0. \quad (22)$$

$$\hat{\boldsymbol{z}}, \hat{\boldsymbol{M}}, \hat{\boldsymbol{l}} = \arg\min_{\boldsymbol{x},\boldsymbol{M},\boldsymbol{l}} \left\{ \sum_k \left\| \boldsymbol{O}_k^l(\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k\boldsymbol{M}_k\boldsymbol{z}) \right\|^2 + \lambda_1 \left\| \boldsymbol{Q}_1\boldsymbol{z} \right\|^2 + \lambda_2 \sum_k \sum_{\boldsymbol{x}} \left\| \boldsymbol{O}_k^h(\boldsymbol{m}_k(\boldsymbol{x}) - \tilde{\boldsymbol{m}}_k(\boldsymbol{x})) \right\|^2 + \lambda_3 \sum_k \left\| \boldsymbol{Q}_2\boldsymbol{m}_k \right\|^2 + \sum_k \sum_{\boldsymbol{x_i}} \sum_{\boldsymbol{x_j} \in \boldsymbol{N}} \boldsymbol{V}_C(l_k(\boldsymbol{x}_i), l_k(\boldsymbol{x}_j)) \right\} \quad (15)$$

Noting that $\boldsymbol{O}_k^l$ and $\boldsymbol{O}_k^h$ are both diagonal matrices, the gradient descent update for the motion field is given by

$$\hat{\boldsymbol{m}}_k^{n+1} = \hat{\boldsymbol{m}}_k^n + \alpha_k^n \left\{ \frac{\partial \boldsymbol{M}_k \boldsymbol{z}}{\partial \boldsymbol{m}_k} \boldsymbol{B}_k^T \boldsymbol{D}_k^T \boldsymbol{O}_k^l (\boldsymbol{y}_k - \boldsymbol{D}\boldsymbol{B}_k \boldsymbol{M}_k \boldsymbol{z}) - \lambda_2 \boldsymbol{O}_k^h (\boldsymbol{m}_k(\boldsymbol{x}) - \tilde{\boldsymbol{m}}_k(\boldsymbol{x})) - \lambda_3 \boldsymbol{Q}_2^T \boldsymbol{Q}_2 \boldsymbol{m}_k \right\} \quad (23)$$

where $\alpha_k^n$ is the step size at the $n$th iteration again, and it must also be selected to be small enough to prevent divergence and large enough to provide faster convergence. As the adaptive selection of $\alpha_k^n$ is beyond the scope of this paper, $\alpha_k^n$ is chosen heuristically. The derivative in (23) is computed on a pixel-by-pixel basis, given by

$$\frac{\partial \boldsymbol{M}_k \boldsymbol{z}(x_u, x_v)}{\partial \boldsymbol{m}_{k,u}} = \frac{z(x_u + m_{k,u} + 1, x_v) - z(x_u + m_{k,u} - 1, x_v)}{2}$$
$$\frac{\partial \boldsymbol{M}_k \boldsymbol{z}(x_u, x_v)}{\partial \boldsymbol{m}_{k,v}} = \frac{z(x_u, x_v + m_{k,v} + 1) - z(x_u, x_v + m_{k,v} - 1)}{2}. \quad (24)$$

Noting that $m_{k,u}$ and $m_{k,v}$ are always fractional values, some interpolation methods are needed to implement the above equations. A bilinear interpolation method is adopted in this paper.

### C. Updating the Segmentation Field

Assuming that HR image $\boldsymbol{z}$ and segment field $\boldsymbol{m}_k$ are known, the cost function to the segment field $\boldsymbol{l}_k$ is given by

$$E_3(\boldsymbol{l}_k) = \lambda_2 \sum_{\boldsymbol{x}} \left\| \boldsymbol{O}_k^h (\boldsymbol{m}_k(\boldsymbol{x}) - \tilde{\boldsymbol{m}}_k(\boldsymbol{x})) \right\|^2 + \sum_{\boldsymbol{x_i}} \sum_{\boldsymbol{x_j} \in \boldsymbol{N}} \boldsymbol{V}_C(l_k(\boldsymbol{x}_i), l_k(\boldsymbol{x}_j)). \quad (25)$$

The segmentation optimization is carried out using the iterated conditional modes (ICM) procedure [38]. The ICM method is a deterministic procedure that aims to reduce the computational load of the stochastic annealing methods. It can best be conceptualized as the "instant freezing" case of the Metropolis algorithm when the temperature is set to zero for all iterations. Similar to simulated annealing, the ICM procedure is implemented at each pixel site separately and cyclically. However, it is in the way by choosing the value at each site that gives the maximum local conditional probability rather than obtaining a value based on the conditional probability distribution; hence, it facilitates faster convergence [33]. This method has been widely applied to the field of image segmentation [34], [39]–[41].

At every iteration step of our ICM segmentation procedure, each pixel is labeled as one of the numbers $1, 2, \ldots R$ under the constraint of minimizing the cost function (25), where $R$ is the number of the independent moving objects. In this process, parametric motion vector $\tilde{\boldsymbol{m}}_k(\boldsymbol{x})$ in (25) is obtained by (3) with the old mapping parameters which are estimated during the last iteration. After the labeling process, these mapping parameters are updated using the least squares estimation according to the current segmentation map. In order to obtain accurate renewed mapping parameters, motion vectors that significantly deviate from the old parameters are not used in the estimation. The available vectors are determined using the following standard:

$$\frac{\left| m_{k,u}^r(\boldsymbol{x}) \right| + \left| m_{k,v}^r(\boldsymbol{x}) \right|}{2} < d \quad (26)$$

where $d$ is a setting threshold.

When the ICM iterations are over, the pixels that were not used to update the mapping parameters in the last step are labeled as 0. These pixels are unobservable because they deviate from all the $K$ motion models. Thus, the diagonal matrix $\boldsymbol{O}_k^h$ can be updated. Once $\boldsymbol{O}_k^h$ is determined, matrix $\boldsymbol{O}_k^l$ can be obtained according to the observation model. Let us review Fig. 1 and assume that the HR pixel H5 has been detected as unobservable. If the optical blur is neglected in the observation model, only the LR pixel L5 which contains H5 is considered unobservable. However, if a $3 \times 3$ blur is considered, the LR pixels L1, L2, L4, and L5, which contain one or more HR pixels of the $3 \times 3$ neighbor of H5, are dependant on H5 according to the observation model. Hence, all the four LR pixels are considered unobservable.

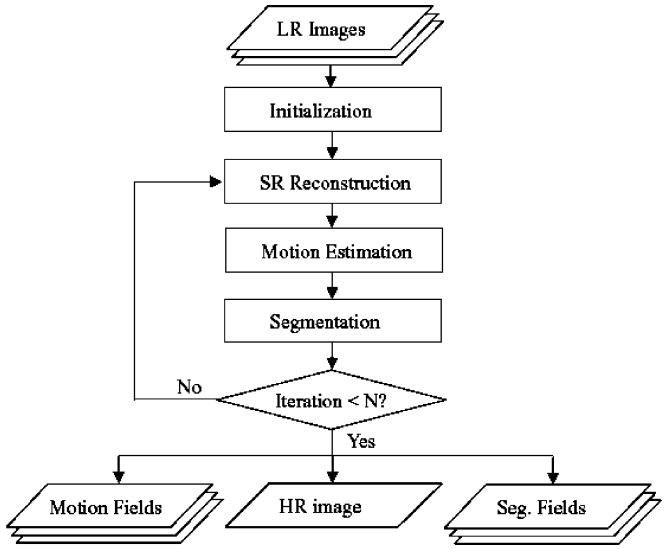


Fig. 2. Diagram of the joint MAP algorithm.

### D. Initialization

Before updating, the HR image, motion fields and segmentation fields must be initiated. The initial HR image is obtained using the bilinear interpolation. The motion estimates are initiated by implementing a blocking matching algorithm, followed by a further estimate using the simple version of (23) by abbreviating the second term in the big bracket. Given the initial motion field, a procedure similar to [42] is adopted to initiate the segmentation. In this procedure, the image is divided into small blocks at first. Then, a set of perspective parameters is estimated for each block. Finally, a k-mean method clusters the candidate perspective parameters in the eight-dimensional parameter space. This results in $R$ regions, each with an initial set of perspective parameters. It is noted that finding the number of segments is still an open and challenging problem. The number,

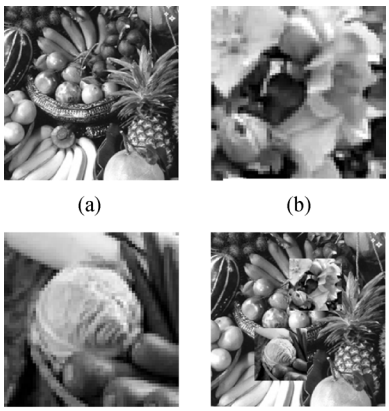

Fig. 3. HR image synthesis in experiment 1: (a) "fruits" image, (b) "flowers" image, (c) "vegetables" image, and (d) the synthetic image.

$R$, is assumed to be known in the initial segmentation. Actually, when the number of moving objects is not very large, we can easily determine it by a visual estimation on the video sequence.

A block diagram of the whole optimization procedure is shown in Fig. 2.

### E. Parameter Determination

Cost function (15) has four parameters $\lambda_1$, $\lambda_2$, $\lambda_3$, and $\gamma$ [see (14)], which control the weights of the second, third, fourth and the fifth terms relative to the first term. The parameter $\lambda_1$ balances the data fidelity (the first term) and the image regularization (the second term) in the reconstruction part. Two different approaches which automatically determine this parameter were used in [21] and [35], respectively. However, it is found $\lambda_1 = 10^{-3}$ is an acceptable choice for most sequences which are not over noisy and smooth in our experiments. Parameter $\lambda_3$ controls the contribution of the motion regularization in the motion estimation part, and we also set it as a fixed value in all the experiments, i.e., $\lambda_3 = 10^4$. The parameter $\lambda_2$ can be chosen as a large value to emphasize the third term if the motion field can be well modeled by the eight-parameter model in (3). It is, otherwise, appropriate to choose a smaller value. As for $\gamma$, Chang *et al.* [34] suggest that it is selected as $1 \leq \lambda_2/\gamma \leq 5$, also depending on how well the motion field conforms to the parametric model. Parameter $\alpha_k^n$ in (23) should be carefully selected because an over small value results in a very slow convergence speed and an over large value often leads to divergence. We set it as $\alpha_k^n = 10^{-6}$ in our experiments. Furthermore, the threshold value ($d$) in (26) must also be set. In the next section, we will demonstrate how it affects the SR results. Our experimental results indicate that $d = 0.25$ is a desirable threshold for most sequences. Lastly, the issue is how to terminate the iterations. Although the proposed algorithm may have different convergences for different image sequence, we can always obtain the desirable results in ten iterations. Therefore, we set $N = 10$ as the termination criterion ($N$ is the iteration number). As the processing of each iteration becomes faster with the increase of the iteration number, this choice will not increase much computation time.

TABLE I
SHIFTS OF THE HR IMAGES (HR PIXEL)

| Image | Attribute | Fruits | | Flowers | | Vegetables | |
|---|---|---|---|---|---|---|---|
| | | H | V | H | V | H | V |
| 0 | Referenced | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | Unreferenced | 0 | 1 | 0 | 3 | 1 | 3 |
| 2 | Unreferenced | 1 | 0 | 1 | 1 | -1 | -2 |
| 3 | Unreferenced | 1 | 1 | 1 | 4 | 3 | 3 |

To sum up this subsection, we have fixed the parameters $\lambda_1$, $\lambda_3$, $\alpha_k^n$, and iteration number $N$, but have chosen different values for parameters $\lambda_2$ and $\gamma$ according to different sequences in our experiments. Our experimental results also indicate that a desirable selection for threshold $d$ is $d = 0.25$.

## V. EXPERIMENTAL RESULTS

Three datasets were employed to illustrate the performance of the proposed algorithm. They are a synthetic image sequence, the "Mobile and Calendar" sequence, and the "Motorcycle and Car" sequence. The following peak signal-to-noise ratio (PSNR) was employed as a quantitative measure in the first two sets of experiments:

$$\text{PSNR} = 10\log_{10}\left(\frac{255^2 * L_1N_1L_2N_2}{\|\hat{\boldsymbol{z}} - \boldsymbol{z}\|^2}\right) \tag{27}$$

where $L_1N_1L_2N_2$ is the total number of pixels in the HR image, and $\hat{\boldsymbol{z}}$ and $\boldsymbol{z}$ represent the reconstructed HR image and the original image, respectively.

### A. Synthetic Image Sequence

In the first set of experiments, three block images, i.e., "fruits," "flowers," and "vegetables," were used to produce the LR data set. The three images are shown in Fig. 3(a)–(c), respectively. The size of "fruits" is $200 \times 200$, and the "flowers" and "vegetables" are both $60 \times 60$. Using the image of "fruits" to simulate the background and regarding "flowers" and "vegetables" as two independent moving objects in the scene, we created four HR images at first. One of the resulting images is shown in Fig. 3(d). In order to simulate the camera pan, not only "flowers" and "vegetables," but also "fruits" were translated to different shifts, which are shown in Table I. After obtaining the HR images, they were down-sampled by a factor of 2 in both horizontal and vertical dimensions to obtain the LR images.

The four resulting LR images were then provided to the proposed algorithm, where $\lambda_2 = 10^4$ and $\gamma = \lambda_2/5 = 2 \times 10^3$ (The determination of other parameters is discussed in Section IV(E)). To test the influence of the threshold $d$ defined in (26) on the estimation process, it was chosen as 0.15, 0.20, 0.25, and 0.30, respectively.

First, the segmentation maps from image 2 are taken as an example for illustration. The interim segmentation maps at the iteration steps 0, 3, 6, and 9 with $d = 0.25$ are shown in Fig. 4(a)–(d), respectively. Many unobservable pixels labeled as "0" exist in the initial segmentation map, as the error of motion estimation is relatively larger. With the increase in iterations, the number of unobservable pixels decreases and further improved segmentation maps are generated. The effect

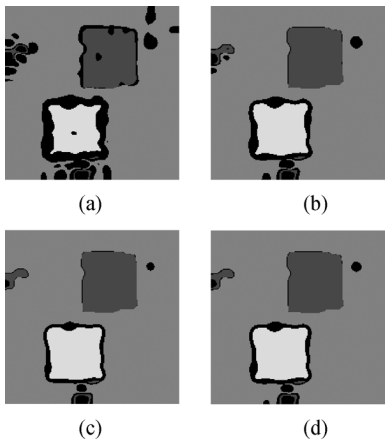


Fig. 4. Interim segmentation maps of image 2 with $d = 0.25$: (a) Iteration $=$ 0, (b) Iteration $= 3$, (c) Iteration $= 6$, and (d) Iteration $= 9$.

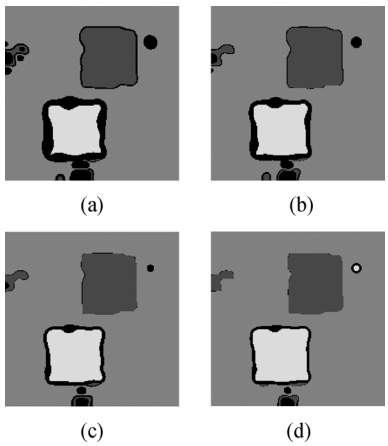


Fig. 5. Segmentation results of image 2 with (a) $d = 0.15$, (b) $d = 0.20$, (c) $d = 0.25$, and (d) $d = 0.30$.

of threshold $d$ to segmentation is demonstrated in Fig. 5. The smaller the value of $d$, greater the number of labeled unobservable pixels in the segmentation map. It is found that this has a direct effect on the motion estimation and SR reconstruction.

Since the real motion vectors are known, the motion estimation was evaluated quantitatively using the mean absolute error (MAE)

$$\text{MAE} = \frac{\sum_{\boldsymbol{x}} |m_u(\boldsymbol{x}) - \hat{m}_u(\boldsymbol{x})| + |m_v(\boldsymbol{x}) - \hat{m}_v(\boldsymbol{x})|}{2L_1N_1L_2N_2} \tag{28}$$

where $\hat{m}_u(\mathbf{x})$ and $\hat{m}_v(\mathbf{x})$ are the estimated parameters, and $m_u(\mathbf{x})$ and $m_v(\mathbf{x})$ are the known values. The MAEs of the three unreferenced images versus the number of iterations are plotted in Fig. 6(a)–(c), respectively. As mentioned previously, the initial motion estimates were obtained using a simple version of (23) without the use of segmentation fields, their MAEs correspond to the first points of the curves. It can be clearly observed that most of the curves are monotonously descending, which proves that the proposed algorithm provides more accurate estimations. Besides, the threshold $d$ significantly influences the motion estimation. In this experiment, the optimal value for motion estimation was 0.25 because it provided accurate and stable results for all the three images.

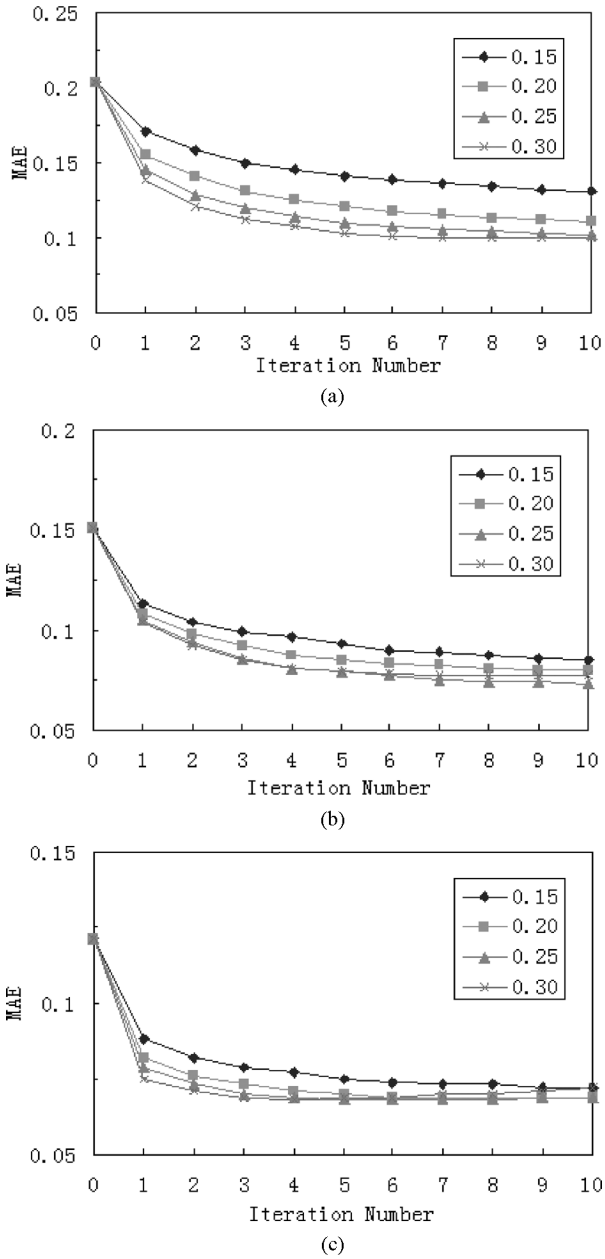


Fig. 6. MAEs versus the iteration numbers with different thresholds $d$, (a) image 1, (b) image 2, and (c) image 3.

Now the emphasis is on the evaluation and analysis of the reconstructed HR images. The referenced LR image is shown in Fig. 7(a). Its bilinear interpolation version is shown in Fig. 7(b), which is viewed as the initial result of the desired HR image in SR reconstruction. The SR result of the conventional MAP algorithm without consideration of the segmentation of unobservable pixels during the reconstruction is shown in Fig. 7(c). Fig. 7(d)–(g) illustrates the reconstruction of the proposed algorithm with different threshold $d$. Detained regions cropped from Fig. 7(c)–(g) and the original image Fig. 3(d) are depicted in Fig. 8(a)–(f), respectively. It is found that the image reconstructed by the conventional MAP algorithm has better visual quality than the interpolated image in most regions. However, artifacts are displayed around the moving object boundaries and/or occlusion regions. The proposed algorithm

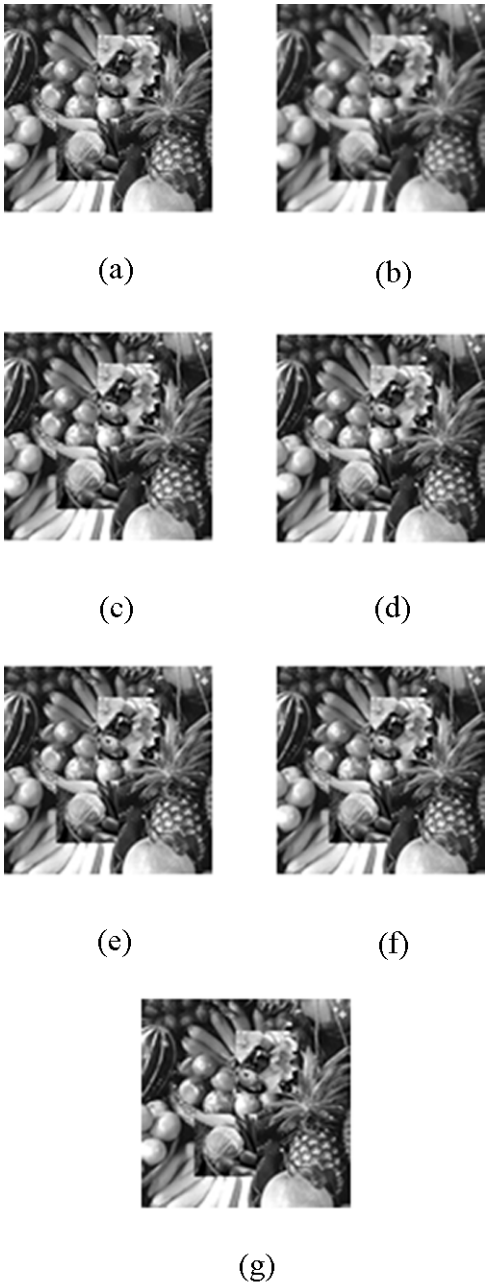


Fig. 7. HR estimates of the synthetic image sequence. (a) The referenced LR image, (b) bilinear interpolation, (c) conventional MAP algorithm and (d)–(g) the proposed algorithm with $d = 0.15$, $d = 0.20$, $d = 0.25$, and $d = 0.30$, respectively.

makes an improvement that it is best described as a suppression of these artifacts and achieves better results under the scenarios of $d = 0.20$ and $d = 0.25$. The two reconstructed results are almost indistinguishable visually. However, when $d = 0.15$, the reconstructed image is too smooth around the motion boundaries as too many pixels are labeled as unobservable and are not considered in the reconstruction. On the other hand, the artifacts are not effectively suppressed when $d = 0.30$ because some real unobservable pixels are not excluded. Quantitative measure agrees with the visual evaluation. PSNR values of Fig. 7(b)–(g) are equal to 24.181, 28.951, 32.020, 32.397, 32.351, and 31.975 dB, respectively. Fig. 9 shows the evolution

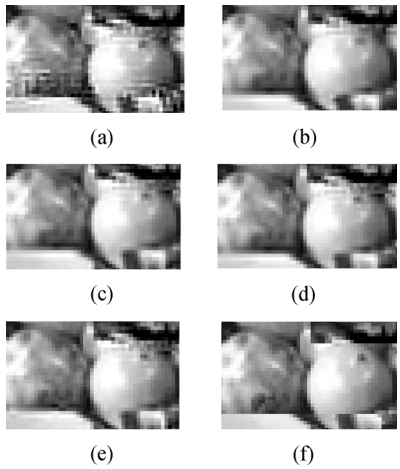


Fig. 8. (a)–(f) Detained regions cropped from Figs. 7(c)–(g) and 3(d), respectively.

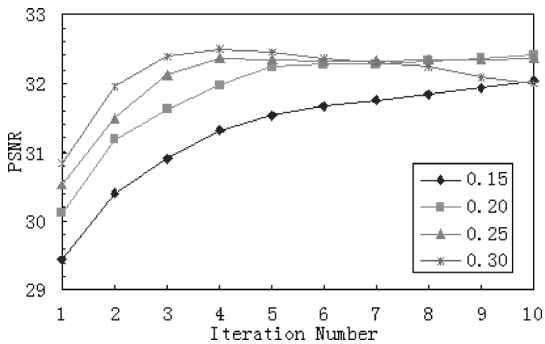


Fig. 9. PSNRs versus the number of iterations in the synthetic image experiment.

of the PSNRs for the proposed algorithm versus the number of iterations. It is apparent that the proposed algorithm has a slowest convergence when $d = 0.15$, and it is very unstable in the case of $d = 0.30$.

### *B. "Mobile and Calendar" Sequence*

In the second set of experiments, the "Mobile and Calendar" sequence was used. These experiments differ from the first set in three ways. First, subpixel motion vectors between image frames are not defined explicitly any longer. They now correspond to the inherent motion within the scene, as introduced by the camera and the objects. Second, the motions are more complicated: the background moves toward the right due to camera pan; the calendar moves up and down; the train moves from right to left; and there is even a revolving ball in the scene. Thirdly, a $3 \times 3$ Gaussian blur was considered in this set of experiments. Although the frame size is $704 \times 576$ pixels in this sequence, our processing was restricted to a typical $352 \times 288$ pixel region to reduce computational expense. The same parameter set was used as in the case of the synthetic sequence except for that $\gamma = \lambda_2 = 10^4$. The threshold $d$ was set to be 0.25.

The initial and final motion estimates of frames 2 and 4 are shown in Fig. 10. It can be found that improvement is obvious after implementing the proposed algorithm. Visual judgment illustrates that the final estimated motion vectors are more consistent with the inherent motions than the initial values. Seg-

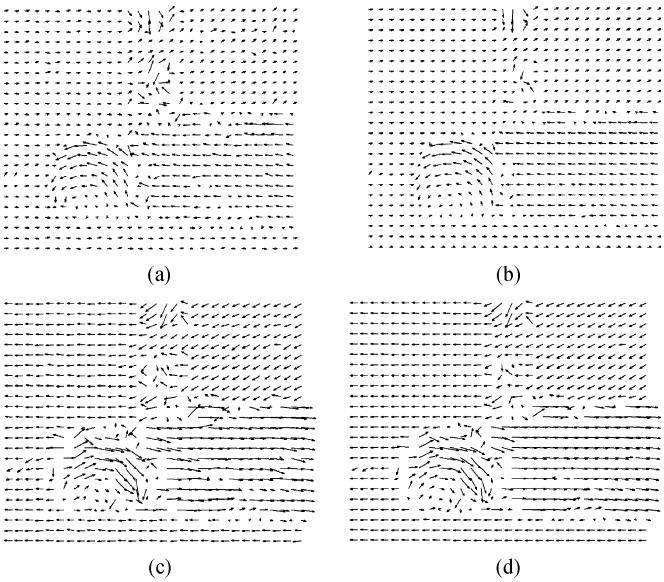

Fig. 10. Motion estimates of the "Mobile and Calendar" sequence. (a) Initial estimates of frame 2, (b) final estimates of frame 2, (c) initial estimates of frame 4, and (d) final estimates of frame 4.

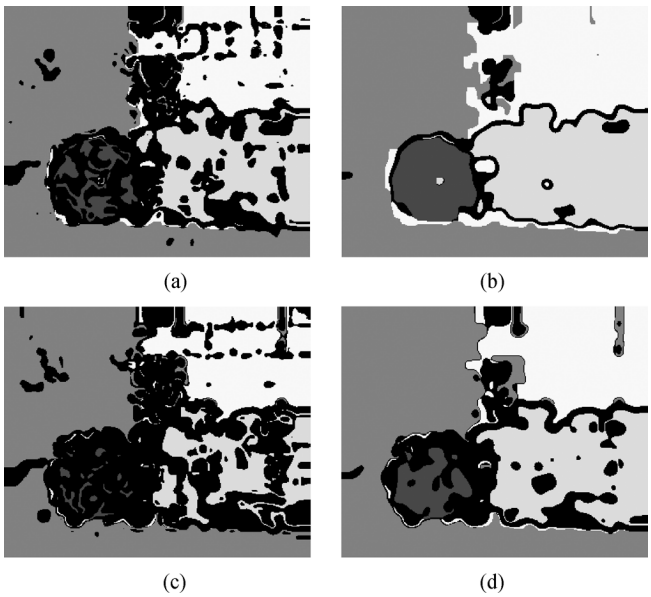

Fig. 11. Segmentation maps of the "Mobile and Calendar" sequence. (a) Initial result of frame 2, (b) final result of frame 2, (c) initial result of frame 4, and (d) final result of frame 4.

mentation maps corresponding to these motion estimates are illustrated in Fig. 11. It can be similarly concluded that in the "Mobile and Calendar" sequence, segmentation maps can also be improved by applying the proposed algorithm. Fig. 12(a)–(d) shows the bilinear interpolated image, the conventional MAP reconstruction result, the result by the proposed cyclic coordinate-descent algorithm, and the original HR image, respectively. Detained regions cropped from Fig. 12(b)–(d) are shown in Fig. 13(a)–(c), respectively. The computed PSNR values for the bilinear, conventional and the proposed methods were equal to 31.079, 33.513, and 34.979, respectively. Evidently, the proposed algorithm outperforms the conventional MAP algorithm

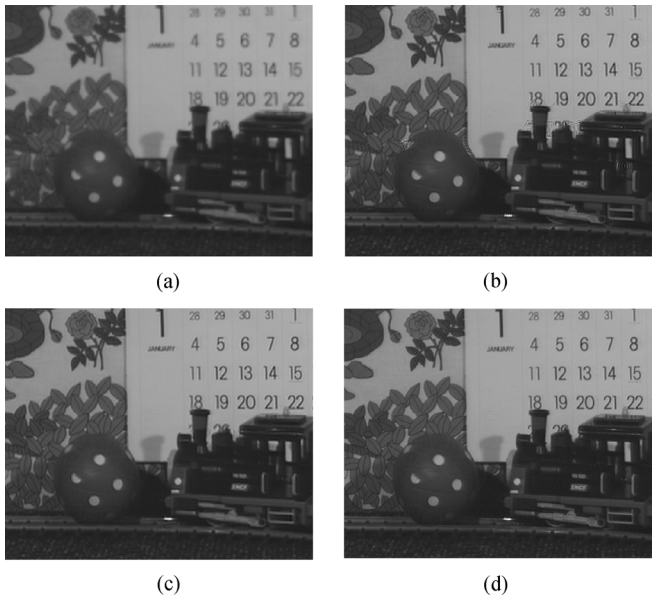

Fig. 12. HR estimates of the "Mobile and Calendar" sequence. (a) Bilinear interpolation, (b) conventional MAP algorithm,(c) the proposed algorithm and (d) the original HR image.

in terms of both the quantitative measurements and visual evaluation. This once again validates the performance of the proposed algorithm, which can suppress artifacts around motion boundaries and occlusion regions. The evolution of the PSNR on this sequence is shown in Fig. 14. Due to the effects of the complicated motions, the relative symmetrical tonality in the image and the considered Gaussian blur, the increase of PSNR in the iteration is not as large as that in synthetic sequence. Nevertheless, it is still a monotone increasing exponential curve.

### C. *"Motorcycle and Car" Sequence*

In each of the two sets of experiments, the sequence is first down sampled, and then the selected referenced frame is reconstructed to the original dimensions using the proposed algorithm. In the third set of experiments, we tested our algorithm on an original sequence, the "Motorcycle and Car," which was captured using a commercial camera. Again, the processing was restricted to a typical $260 \times 130$ pixel region to reduce computational cost. We used five frames with frame 3 (the referenced frame) which is shown in Fig. 15(a). The parameters was set as: $\lambda_2 = 10^3$, and $\gamma = (1/5)\lambda_2 = 2 \times 10^2$ for frame 2 and frame 4 and $\gamma = \lambda_2 = 10^3$ for other two frames. The threshold $d$ was set to be 0.25 again. Fig. 15(b)–(d) shows the bilinear interpolated image, the conventional MAP reconstruction result, and the result by the proposed algorithm, respectively. It is apparent the proposed algorithm obtained more desirable result than the bilinear interpolation and the conventional MAP algorithm. The motion field and segmentation field of frame 2 are, respectively, shown in Figs. 16 and 17 as illustrative examples. To illustrate the convergence behavior of the estimator, the MAP cost function in (15) is plotted in Fig. 18 against the iteration number.

## VI. Conclusion

Reconstructing an HR image from several LR images using SR techniques becomes complicated when the scenes contain


Fig. 13. (a)–(c) Detained regions cropped from Fig. 12(b)–(d), respectively.

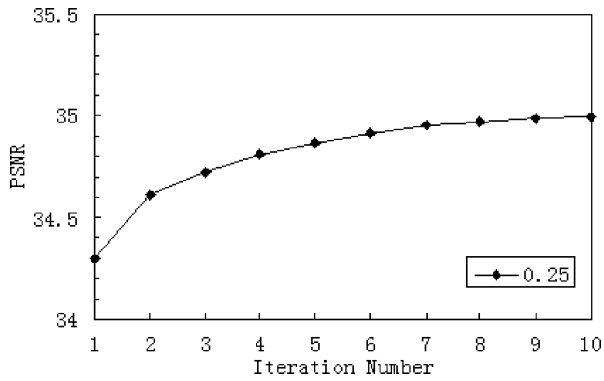


Fig. 14. PSNR versus the number of iterations in the experiment of "Mobile and Calendar" sequence.

multiple independently moving objects. To address this challenge, we proposed a joint MAP formulation combining motion estimation, segmentation, and SR together. The formulation is solved by a cyclic coordinate decent process that treats the motion fields, segmentation fields, and HR image as unknowns and estimates them jointly using the available data. The advantage of this algorithm is that the motion estimates, segmentation maps and HR image can benefit each other. The proposed

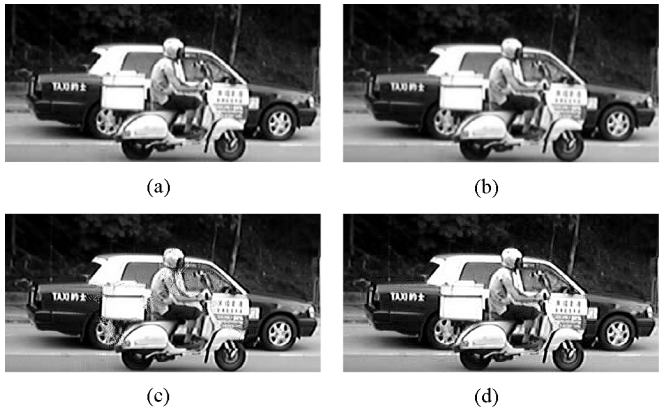


Fig. 15. HR estimates of the "Motorcycle and Car" sequence. (a) Original LR image (frame 3), (b) bilinear interpolation, (c) conventional MAP algorithm, and (d) the proposed algorithm.

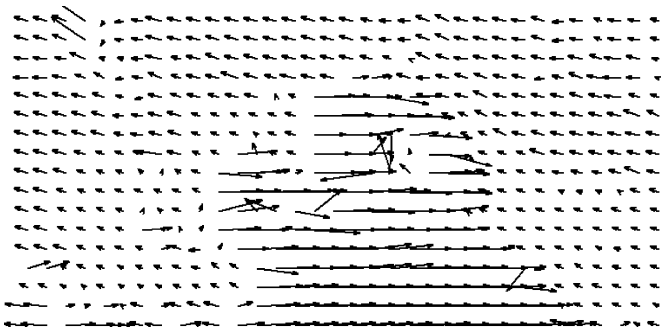

Fig. 16. Motion estimates of frame 2 of the "Motorcycle and Car" sequence.

Fig. 17. Segmentation map of frame 2 of the "Motorcycle and Car" sequence.

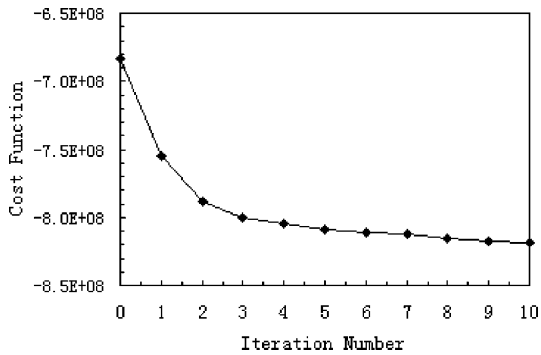


Fig. 18. Cost function value versus the number of iterations in the experiment of the "Motorcycle and Car" sequence.

algorithm was tested on a synthetic image sequence, the "Mobile and Calendar" sequence and the original "Motorcycle and

Car" sequence. Experiment results validated that motion estimates and segmentation maps can be noticeably improved by implementing this algorithm. Moreover, the proposed algorithm suppresses artifacts around motion boundaries and occlusion regions and provides more desirable reconstructed HR image when compared with the conventional algorithm. Nevertheless, there may still be room for the improvement of our optimal method to increase the computational efficiency. Using more robust regularizations for the image and/or the motion field—e.g., Huber [15], total variance (TV) [43], and bilateral-TV [44], [45]—may further improve the SR results. These will be addressed in our future work.

## REFERENCES


[1] R. Y. Tsai and T. S. Huang, "Multi-frame image restoration and registration," *Adv. Comput. Vis. Image Process.*, vol. 1, pp. 317–339, 1984.

[2] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 2, pp. 1013–1027, Jun. 1990.

[3] S. P. Kim and W. Y. Su, "Recursive high-resolution reconstruction of blurred multiframe images," *IEEE Trans. Image Process.*, vol. 2, no. 4, pp. 534–539, Apr. 1993.

[4] S. Rhee and M. G. Kang, "Discrete cosine transform based regularized high-resolution image reconstruction algorithm," *Opt. Eng.*, vol. 38, pp. 1348–1356, 1999.

[5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.

[6] H. Ur and D. Gross, "Improved resolution from sub-pixel shifted pictures," *CVGIP: Graph. Models Image Process.*, vol. 54, pp. 181–186, 1992.

[7] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circuits Syst.*, vol. CAS-24, no. 11, pp. 652–654, Nov. 1977.

[8] L. J. Yen, "On nonuniform sampling of bandwidth limited signals," *IRE Trans. Circuits Theory*, vol. 3, pp. 251–257, 1956.

[9] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graph. Models Image Process.*, vol. 53, pp. 231–239, 1991.

[10] H. Stark and P. Oskoui, "High-resolution image recovery from image plane arrays, using convex projections," *J. Opt. Soc. Amer. A: Opt. Image Sci., Vis.*, vol. 6, pp. 1715–1726, 1989.

[11] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing, San Francisco, CA, 1992.

[12] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "High-resolution image reconstruction from a low-resolution image sequence in the presence of time-varying motion blur," presented at the IEEE Int. Conf. Image Processing, Austin, TX, 1994.

[13] ——, "Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1064–1076, Aug. 1997.

[14] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image from multiple-degraded misregistered low-resolution images," *Proc. SPIE*, 1994.

[15] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Process.*, vol. 5, no. 6, pp. 996–1011, Jun. 1996.

[16] R. C. Hardie, T. R. Tuinstra, J. Bognar, K. J. Barnard, and E. Armstrong, "High resolution image reconstruction from digital video with global and non-global scene motion," presented at the IEEE Int. Conf. Image Processing, Santa Barbara, CA, 1997.

[17] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1646–1658, Dec. 1997.

[18] ——, "Superresolution restoration of an image sequence: Adaptive filtering approach," *IEEE Trans. Image Process.*, vol. 8, no. 3, pp. 387–395, Mar. 1999.

[19] N. K. Bose, H. C. Kim, and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," presented at the IEEE Int. Conf. Acoustics, Speech and Signal Processing, Minneapolis, MN, 1993.

[20] M. K. Ng and N. K. Bose, "Analysis of displacement errors in high-resolution image reconstruction with multisensors," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 49, no. 6, pp. 806–813, Jul. 2002.

[21] E. S. Lee and M. G. Kang, "Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 826–837, Jul. 2003.

[22] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Trans. Image Process.*, vol. 8, no. 8, pp. 879–885, Aug. 1999.

[23] B. C. Tom and A. K. Katsaggelos, "Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images," presented at the IEEE Int. Conf. Image Processing, Washington, DC, 1995.

[24] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Process.*, vol. 6, no. 12, pp. 1621–1633, Dec. 1997.

[25] C. A. Segall, R. Molina, and A. K. Katsaggelos, "High-resolution images from low-resolution compressed video," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 37–48, May 2003.

[26] C. A. Segall, A. K. Katsaggelos, R. Molina, and J. Mateos, "Bayesian resolution enhancement of compressed video," *IEEE Trans. Image Process.*, vol. 13, no. 7, pp. 898–910, Jul. 2004.

[27] N. A. Woods, N. P. Galatsanos, and A. K. Katsaggelos, "Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 201–213, Jan. 2006.

[28] J. Chung, E. Haber, and J. Nagy, "Numerical methods for coupled super resolution," *Inv. Probl.*, to be published.

[29] S. Borman and R. Stevenson, Spatial Resolution Enhancement of Low-Resolution Image Sequences: A Comprehensive Review with Directions for Future Research Lab. Image and Signal Analysis (LISA), Univ. Notre Dame, Notre Dame, IN, 1998.

[30] M. Irani and S. Peleg, "Motion analysis for image enhancement. Resolution, occlusion, and transparency," *J. Vis. Commun. Image Represent.*, vol. 4, p. 324, 1993.

[31] P. E. Eren, M. I. Sezan, and A. M. Tekalp, "Robust, object-based high-resolution image reconstruction from low-resolution video," *IEEE Trans. Image Process.*, vol. 6, no. 10, pp. 1446–1451, Oct. 1997.

[32] L. D. Alvarez, J. Mateos, R. Molina, and A. K. Katsaggelos, "High-resolution images from compressed low-resolution video: Motion estimation and observable pixels," *Int. J. Imag. Syst. Technol.*, vol. 14, pp. 58–66, 2004.

[33] A. M. Tekalp, *Digital Video Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[34] M. M. Chang, A. M. Tekalp, and M. I. Sezan, "Simultaneous motion estimation and segmentation," *IEEE Trans. Image Process.*, vol. 6, no. 9, pp. 1326–1333, Sep. 1997.

[35] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Trans. Image Process.*, vol. 10, no. 9, pp. 1299–1308, Sep. 2001.

[36] S. Hsu, P. Anandan, and S. Peleg, "Accurate computation of optical flow by using layered motion representations," presented at the Int. Conf. Pattern Recognition, Jerusalem, Israel, 1994.

[37] A. J. Tabatabai, R. S. Jasinschi, and T. Naveen, "Motion estimation methods for video compression – A review," *J. Franklin Inst. Eng. Appl. Math. B*, vol. 335, pp. 1411–1441, 1998.

[38] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc. B*, pp. 259–302, 1986.

[39] M. M. Chang, M. I. Sezan, and A. M. Tekalp, "An algorithm for simultaneous motion estimation and scene segmentation," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing, 1994.

[40] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 901–914, Apr. 1992.

[41] J. K. Fwu and P. M. Djuric, "Unsupervised vector image segmentation by a tree structure – ICM algorithm," *IEEE Trans. Med. Imag.*, vol. 15, no. 6, pp. 871–880, Dec. 1996.

[42] J. Y. A. Wang and E. H. Adelson, "Representing moving images with layers," *IEEE Trans. Image Process.*, vol. 3, no. 5, pp. 625–638, Sep. 1994.

[43] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," presented at the 8th Eur. Conf. Computer Vision, Prague, Czech Republic, 2004.

[44] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.

[45] S. Farsiu, M. Elad, and P. Milanfar, "Multiframe demosaicing and super-resolution of color images," *IEEE Trans. Image Process.*, vol. 15, no. 1, pp. 141–159, Jan. 2006.

**Huanfeng Shen** received the B.S. degree in surveying and mapping engineering from Wuhan University, Wuhan, China, in 2002, where he is currently pursuing the Ph.D. degree at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

His current research interests focus on image reconstruction, remote sensing, image processing, and application.

**Liangpei Zhang** received the B.S. degree in physics from the Hunan Normal University, ChangSha, China, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1982, 1988, and 1998, respectively.

From 1997 to 2000, he was a Professor in the School of the Land Sciences, Wuhan University. In August 2000, he joined the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, as a Professor and Head of the Remote Sensing Section. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang has served as Co-Chair of the SPIE Series Conferences on Multispectral Image Processing and Pattern Recognition (MIPPR) and the Conference on Asia Remote Sensing in 1999, Editor of the MIPPR01, MIPPR05 Symposia, and the Chinese National Committee for the International Geosphere-Biosphere Programme. He is an Associate Editor of the *Geo-Spatial Information Science Journal*

**Bo Huang** is currently an Associate Professor in the Department of Geography and Resource Management, The Chinese University of Hong Kong (CUHK), and Deputy Director of the Joint Laboratory of GeoInformation Science, CUHK, and Chinese Academy of Sciences. He also holds Adjunct Professorships at the University of Calgary, Calgary, AB, Canada, and several universities and research institutes in China. His research interests are in the areas of geographic information systems, image processing, spatial statistics, and spatial optimization. He has (co-)authored more than 40 articles in international journals.

**Pingxiang Li** received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University,Wuhan, China, in 1986, 1994, and 2003, respectively.

Since 2002, he has been a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. His research interests include photogrammetry and SAR image processing.