A super-resolution reconstruction algorithm for surveillance images

Liangpei Zhang\textsuperscript{a,}\textsuperscript{*}, Hongyan Zhang\textsuperscript{a}, Huanfeng Shen\textsuperscript{b}, Pingxiang Li\textsuperscript{a}

\textsuperscript{a} The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, China
\textsuperscript{b} School of Resource and Environmental Science, Wuhan University, Wuhan, Hubei, China

\textbf{Abstract}

In many surveillance video applications, it is of interest to recognize a region of interest (ROI), which often occupies a small portion of a low-resolution, noisy video. This paper proposes an edge-preserving maximum a posteriori (MAP) estimation based super-resolution algorithm using a weighted directional Markov image prior model for a ROI from more than one low-resolution surveillance image. Conjugate gradient (CG) optimization based on standard operations on images is then developed to improve the computational efficiency of the algorithm. The proposed algorithm is tested on different series of surveillance images. The experimental results indicate that the proposed algorithm has considerable effectiveness in terms of both objective measurements and visual evaluation.

© 2009 Elsevier B.V. All rights reserved.

\textbf{1. Introduction}

High-resolution (HR) images are useful in many applications such as surveillance, video frame freezing, medical diagnostics, and military information gathering, etc. However, because of the high cost and physical limitations of the high precision optics and image sensors, it is not easy to obtain the desired HR images in many cases. Therefore, super-resolution (SR) image reconstruction techniques, which can reconstruct one or a set of HR images from a sequence of low-resolution (LR) images of the same scene, have been widely researched in the last two decades.

Super-resolution image reconstruction refers to a process that produces a HR image from a sequence of LR images. It overcomes the inherent resolution limitation by bringing together the additional information from each image. Generally, SR techniques can be divided into frequency domain algorithms and spatial domain algorithms. Much of the earlier SR work was developed in the frequency domain using discrete Fourier transform (DFT), such as the work of Tsai and Huang [1], Kim et al. [2,3] and so on. More recently, discrete cosine transform (DCT) [4] and wavelet transform based SR methods [5–7] have also been proposed. In the spatial domain, typical reconstruction models include non-uniform interpolation [8], iterative back projection (IBP) [9], projection onto convex sets (POCS) [10–13], maximum likelihood (ML) [14], maximum a posteriori (MAP) [15–17], hybrid ML/MAP/POCS [18], and adaptive filtering [19]. Based on these basic reconstruction models, researchers have produced many extended algorithms, such as the joint formulation of reconstruction and registration [20–23], algorithms for multi-spectral and color [24,25], hyper-spectral [26], and compressed [27,28] sequence, etc. Amongst the numerous solution techniques featuring in the literature, the MAP estimation method is one of the most promising.

In many surveillance video applications, it is of interest to recognize an object that is selected as a region of interest (ROI), such as numbers, words, or labels that often occupy a small portion of a low-resolution, noisy video [29]. In these circumstances, edges of the object are often very useful and important for surveillance applications, but the commonly used Gaussian
Markov random field regularization (GMRF) in the MAP-based SR cannot effectively preserve sharp edges in the estimated images. To achieve this, several edge-preserving regularizations have been proposed, such as Huber–Markov regularization [15] and bilateral-TV regularization [30]. This paper proposes a weighted directional Markov image prior model, which utilizes different weights for different directional smoothness measures of the edge pixels to reduce the smoothing effect of the common Markov random field prior along edges. Conjugate gradient optimization is then developed to estimate the HR image. Standard operations on images such as convolution, warping, sampling are used to improve the computational efficiency of the algorithm. Experimental results show that the proposed algorithm can effectively preserve edges and detailed information in the reconstructed image.

The rest of the paper is organized as follows. Section 2 presents the image observation model of the SR problem. The motion estimation method employed is described in Section 3. In Section 4, the MAP estimation based reconstruction algorithm is explained. Experimental results are provided in Section 5, and Section 6 concludes this paper.

2. Image observation model

The image observation model is employed to relate the desired referenced HR image to all the observed LR images. Typically, the imaging process involves warping, followed by blurring and down-sampling to generate LR images from the HR image. Let the underlying HR image be denoted in the vector form by \( z = [z_1, z_2, \ldots, z_{L_1 N_1 L_2 N_2}] \), where \( L_1 N_1 \times L_2 N_2 \) 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 has the size \( N_1 \times N_2 \). Thus, the LR image can be represented as \( y_k = [y_{k,1}, y_{k,2}, \ldots, y_{k,N_1 N_2}] \), where \( k = 1, 2, \ldots, P \), with \( P \) being the number of LR images. Assuming that each observed image is contaminated by additive noise, the observation model can be represented as [18]

\[
y_k = DB_k M_k z + n_k
\]

where \( M_k \) is the warp matrix with the size of \( L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2 \), \( B_k \) represents the camera blur matrix also of size \( L_1 N_1 L_2 N_2 \times L_1 N_1 L_2 N_2 \), \( D \) is a \( N_1 N_2 \times N_1 N_2 \) down-sampling matrix, and \( n_k \) represents the \( N_1 N_2 \times 1 \) noise vector. It should be noted that all the images are assumed to have the same blurring function, so the matrix \( B_k \) can be substituted by \( B \)

\[
y_k = DB_k M_k z + n_k
\]

This paper proposes a weighted directional Markov image prior model, which utilizes different weights for different directional smoothness measures of the edge pixels to reduce the smoothing effect of the common Markov random field prior along edges. Conjugate gradient optimization is then developed to estimate the HR image. Standard operations on images such as convolution, warping, sampling are used to improve the computational efficiency of the algorithm. Experimental results show that the proposed algorithm can effectively preserve edges and detailed information in the reconstructed image.

3. Motion estimation method

Motion estimation/registration plays a critical role in SR reconstruction, where it is necessary to select a frame from the sequence as the reference one. In general, the sub-pixel motions between the reference frame and the unrelated frame can be modeled and estimated by a parameter model. This section introduces the motion estimation methods employed.

It is assumed that the motion of the ROI during the sequence is a globally translational motion and the motions of all points can often be modeled by a parametric model. Generally, the relationship between the observed \( k \)th and \( l \)th frames can be expressed by

\[
y_k(x_u, x_v) = y_l^{(l)}(x_u, x_v) + e_k(x_u, x_v)
\]

where \((x_u, x_v)\) denotes the pixel site, \( y_k(x_u, x_v) \) is a pixel in frame \( k \), \( \theta \) is the vector containing the corresponding motion parameters, \( y_l^{(l)}(x_u, x_v) \) is the predicted pixel of \( y_k(x_u, x_v) \) from frame \( l \) using parameter vector \( \theta \), and \( e_k(x_u, x_v) \) denotes the model error. In the literature, the six-parameter affine model and eight-parameter perspective model are widely used. Here the concentration is on the affine model, in which \( y_k^{(l)}(x_u, x_v) \) can be expressed as

\[
y_k^{(l)}(x_u, x_v) = y_l(a_0 + a_1 x_u + a_2 x_v, b_0 + b_1 x_u + b_2 x_v)
\]

In this model, \( \theta = (a_0, a_1, a_2, b_0, b_1, b_2)^T \) contains six geometric model parameters. To solve \( \theta \), we can employ the least square criteria, which has the following minimization cost function:

\[
E(\theta) = \|y_k - y_k^{(l)}\|^2
\]

Using the Gauss–Newton method, the six affine parameters can be iteratively solved by

\[
\Delta \theta = (J^T J)^{-1} [-J^T e]
\]

Fig. 1 illustrates this equation.

![Fig. 1. Block diagram illustration of the observation model (2), where the desired high-resolution image is at the extreme left with the observed image at the extreme right.](image-url)
and
\[ \theta^{n+1} = \theta^n + \Delta \theta \] (7)
Here, n is the iteration number, \( \Delta \theta \) denotes the corrections of the model parameters, \( \theta^n \) is the residual vector that is equal to \( y_k - y_k^{(0/n)} \), and \( J^n = \partial \theta^n / \partial \theta \) denotes the gradient matrix of \( \theta^n \).

4. Image reconstruction algorithm

4.1. MAP formulation

In most situations, the problem of SR is an ill-posed inverse one because the information contained in the observed LR images is not sufficient to solve the HR image. In order to obtain more desirable SR results, the ill-posed problem should be stabilized to be well-posed. The MAP method, which can easily include image prior or regularization, is a good framework to describe the SR problem from the statistical point of view. Let the full set of P LR images be denoted by \( y = [y_1, y_2, \ldots, y_P]^T \). The purpose is to realize the MAP estimate of HR image \( z \), given the observed LR images \( y \). The estimate can be computed by

\[ \hat{z} = \arg \max p(z|y) \] (8)

Applying Bayes’ rule, Eq. (8) becomes

\[ \hat{z} = \arg \max \frac{p(y|z)p(z)}{p(y)} \] (9)

Since \( p(y) \) can be considered a constant and can be eliminated from the optimization in Eq. (9), it can be rewritten as

\[ \hat{z} = \arg \max p(y|z)p(z) \] (10)

Assuming the LR images are independent, we obtain

\[ \hat{z} = \arg \max \prod_{k} p(y_k|z)p(z) \] (11)

Using the monotonic logarithm function, it can be expressed as

\[ \hat{z} = \arg \max \sum_{k} \log p(y_k|z) + \log p(z) \] (12)

Usually, the noise \( n_k \) in (1) is assumed to be one of the zero mean white Gaussian types; thus the likelihood distribution \( p(y_k|z) \) has the form

\[ p(y_k|z) = \frac{1}{C_1} \exp \left( -\frac{||y_k - A_k z||^2}{2\sigma_k^2} \right) \] (13)

where \( C_1 \) is a constant, and \( \sigma_k^2 \) is the error variance. The image prior \( p(z) \) in (12) has the Gibbs form

\[ p(z) = \frac{1}{C_2} \exp(-\Gamma(z)) \] (14)

where \( C_2 \) is also a constant, and \( \Gamma(z) \) is the prior energy functional. Substituting (13) and (14) in (12), after some manipulation, \( C_1, C_2, \) and \( \sigma_k^2 \) can be safely dropped, and the maximization of this posterior probability distribution is equivalent to the following regularization problem:

\[ \hat{z} = \arg \min \sum_{k} ||y_k - A_k z||^2 + \lambda \Gamma(z) \] (15)

where the first term \( \sum_{k} ||y_k - A_k z||^2 \) is the data fidelity term, \( \Gamma(z) \) acts as the regularization term, and \( \lambda \) is the regularization parameter.

4.2. Weighted directional Markov regularization

For the regularization term, Tikhonov regularization [16,18,21,27,28] and Gauss–Markov [20] are commonly employed. A common criticism of these regularization methods is that the sharp edges and detailed information in the estimates tend to be overly smoothed [31]. Here, a weighted directional Markov regularization is proposed to preserve the edges and detailed information in the reconstruction process.

The energy functional of the common Markov random field is defined as

\[ \Gamma(z) = \sum_{c \in C} V_c(z) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \sum_{m=1}^{4} w_m \rho(d_{ij}^m z) \] (16)

where \( V_c(z) \) is some function of a local group of points \( c \) called cliques, and \( C \) denotes the set of all cliques throughout the image. \( \rho(\cdot) \) is the function of smooth measure \( d_{ij}^m z \), \( \rho(\cdot) \) can use a quadratic function or a Huber function, which correspond to Gauss–Markov regularization and Huber–Markov, respectively.

\[ d_{ij}^m z \] denotes the directional smooth measures at pixel \( z_{ij} \) and they are given as

\[ d_{ij}^1 x = x_{ij+1} - 2x_{ij} + x_{ij-1} \]

\[ d_{ij}^2 x = \frac{\sqrt{2}}{2}(x_{i-1,j+1} - 2x_{ij} + x_{i+1,j+1}) \]

\[ d_{ij}^3 x = x_{i+1,j} - 2x_{ij} + x_{i-1,j} \]

\[ d_{ij}^4 x = \frac{\sqrt{2}}{2}(x_{i-1,j+1} - 2x_{ij} + x_{i+1,j-1}) \] (17)

It is noted that the intensity difference between the pixels perpendicular to the image edges is larger, while that between the pixels along the edges is smaller. The traditional MRF regularization poses the same constraint coefficients for the four directional smooth measures, with the same extent constraint to the central edge pixel from the neighborhood pixels, which causes the sharp edges and detailed information in the estimated image to be overly smoothed. To preserve the sharp edges in the estimate, the constraint between the pixels perpendicular to the edges should be reduced and that between the pixels along the edges should be increased. For this, the weighted Markov random field (WMRF) regularization is proposed to preserve the sharp edges.

The energy functional of the WMRF is defined as

\[ \Gamma(z) = \sum_{c \in C} V_c(z) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \sum_{m=1}^{4} w_m \rho(d_{ij}^m z) \] (18)

where \( w_m \) (\( m = 1,2,3,4 \)) are the positive weight coefficients for the directional smooth measure.

The Prewit operators are first used to detect edge pixels in the image. For the non-edge pixels, the four weights are set to \( w_1 = w_2 = w_3 = w_4 = 1 \); while for the edge pixels,
the weights should satisfy the following properties: 
(a) the weights \( w_k \) should be inversely proportional to the smooth measure function \( \rho(d_{ij}^2) \); 
(b) \( \sum_{m=1}^{4} w_m = 4 \). 
This constraint avoids the trivial solution (all zeros) for \( w_m \). The simplest linear solution for criteria (a) and (b) is

\[
  w_m = \frac{R_{\text{ave}}}{\rho(d_{ij}^2) + \beta}
\]

(19)

where \( \beta \) is a small positive parameter, which prevents the denominator from becoming zero. \( R_{\text{ave}} \) is defined as

\[
  R_{\text{ave}} = \frac{4}{\sum_{m=1}^{4} \rho(d_{ij}^2) + \beta}
\]

(20)

\( w_m \) depends on the smooth measure at pixel \( z_{ij} \).

---

**Fig. 2.** Simulation results of deblurring using different regularizations. (a) Original image. (b) Blurred and noisy image. (c) Deblurring result with GMRF regularization. (d) Deblurring result with GWMRF regularization. (e) Deblurring result with HMRF regularization. (f) Deblurring result with HWMRF regularization.
4.3. Numerical solution

It is time-consuming to solve the super-resolution problem in the form of matrix-vector because of the very high dimensionality of vector $z$. A considerable speedup in the super-resolution computation can be achieved by taking advantage of the following two methods: implementing advanced gradient based optimization techniques and computing the gradient in an efficient manner.

In this paper, we derive the conjugate gradient (CG) optimization method to solve the cost function (15). The CG optimization utilizes conjugate direction instead of local gradient to search for the minima. Therefore, it can achieve faster convergence when compared to the steepest descent method [32]. It also requires less storage requirements and computation complexity compared with the Quasi-Newton method.

The gradient of the cost function is denoted as $r(z^n) = \sum_k A_k^T (A_k z^n - y_k) + \lambda \nabla I(z^n)$. The right term of the gradient $\nabla I(z^n)$ is the derivative of the regularization term with respect to $z$ and can be approximated from the estimated HR image. While the left term $A_k^T (A_k z^n - y_k) = M_k B^T D^T (DBM_k z^n - y_k)$ can be computed using basic image operations such as warp, blur, and sampling instead of computational intensive matrix-vector multiplication.

Fig. 3. (a–f) Detail regions cropped from Fig. 2(a–f), respectively.
sparse matrices multiplications. The matrices $M_k$, $B$, $D$ model the image formation process, and their implementation is simply the image warping, blurring, and downsampling, respectively. The implementation of their transpose matrices is also very simple. $D^T$ is implemented by up-sampling the image without interpolation, i.e., by zero padding. For a convolution blur, $B^T$ is implemented by convolution with the flipped kernel of the imaging blur.

Table 1
The MSE errors of super-resolution reconstruction using different regularizations.

<table>
<thead>
<tr>
<th></th>
<th>GMRF</th>
<th>GWMRF</th>
<th>HMRF</th>
<th>HWMRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>First experiment</td>
<td>18.79</td>
<td>14.61</td>
<td>12.91</td>
<td>9.73</td>
</tr>
<tr>
<td>Second experiment</td>
<td>77.54</td>
<td>72.02</td>
<td>71.85</td>
<td>62.51</td>
</tr>
</tbody>
</table>

Fig. 4. Results of super-resolution reconstruction using different regularizations. (a) Original image. (b) LR image. (c) Reconstruction result with GMRF regularization. (d) Reconstruction result with GWMRF regularization. (e) Reconstruction result with HMRF regularization. (f) Reconstruction result with HWMRF regularization.
kernel \( b(i,j) \). If \( M_k \) is implemented by backward warping, then \( M^T_k \) should be the forward warping of the inverse motion \([33]\). Thus, the gradient of the cost function with the WMRF regularization is computed in an efficient manner and the CG optimization technique can be used without explicit construction of these large matrices.

5. Experiments

To test the performance of the proposed algorithm, a number of experiments were conducted, four of which are presented here. The first two were controlled simulated experiments, and the last two were real surveillance image sequences with unknown motion vectors. The following mean square error (MSE) \([34]\) was employed as the quantitative measure:

\[
MSE = \frac{||z - \hat{z}||^2}{L_1N_1 \times L_2N_2}
\]

where \( L_1N_1L_2N_2 \) is the total number of pixels in the HR image, \( \hat{z} \) and \( z \) represent the reconstructed HR image and the original image, respectively. Here the quadratic

![Fig. 5. (a–e) Detailed regions cropped from Fig. 4(a, c–f), respectively.](image-url)
function and the Huber function are both chosen as the function of smooth measure to compare the performance of MRF and WMRF regularizations.

5.1. Controlled simulated experiments

In the first experiment, the superiority of the WMRF regularization over MRF regularization was justified. We corrupted an image by blurring it with a $3 \times 3$ Gaussian blur kernel with standard deviation equal to 0.5. Gaussian additive noise was added to the resulting blurred image to achieve SNR equal to 25 dB. If $z$ and $y$ represent the original and corrupted images, and $B$ represents the matrix of blur kernel, the blurred noisy image can be reconstructed by minimizing the following:

$$
\hat{z} = \arg \min \{ |y - Bz|^2 + \lambda J(z) \}
$$

(23)

The criterion for regularization parameter selection in this experiment was to choose parameters that produced the most visually appealing results. Therefore, to ensure fairness, each method was repeated several times with different regularization parameters and the best result was chosen as the output of the method [30].

Figs. 2(a) and 3(a) show the original image and its zoom. Figs. 2(b) and 3(b) are the corrupted blue + noise, where $n$ is the Gaussian additive noise. The image was reconstructed using MRF and WMRF regularizations, and this scenario can be thought of as a super-resolution reconstruction problem with a resolution enhancement factor of one.

The reconstruction results are shown in Figs. 2 and 3. The MSE errors of these results are listed in Table 1. The selected threshold for Prewit operators is 35. When the quadratic function is chosen as the function of the smooth measure, the reconstruction result of GMRF regularization is shown in Figs. 2(c), 3(c) and achieves the smallest MSE error of 18.79; while Figs. 2(d) and 3(d) are the reconstruction result of the proposed GWMRF regularization, which has the best MSE value of 14.61. When using the Huber function as the function of the smooth measure, the reconstruction result of the edge-preserving HMRF regularization proposed in [15] is shown in Figs. 2(e), 3(e) and has the smallest MSE error of 12.91; the proposed HWMRF regularization achieves the MSE error of 9.73, and the reconstructed image is shown in Figs. 2(f) and 3(f). From the comparison of Fig. 3(c–f), the reconstruction result of GWMRF regularization is better than GMRF and close to HMRF, while HWMRF achieves the best result. We notice that whichever function is chosen as the function of the smooth measure, the proposed WMRF regularization achieves better results than MRF regularization in terms of both the quantitative measurements and visual evaluation.

5.2. Surveillance sequences

Here we use real surveillance sequences to validate the effectiveness of the proposed super-resolution method. It is assumed that the motion of the images during the sequence is a globally translational motion; the motions between the reference image and the unreference images can be estimated by the affine parameter model introduced in Section 3.

The first surveillance sequence experiment shows a SR reconstruction of a car sequence, which was obtained from a surveillance video camera. One frame of this sequence is shown in Fig. 6. One frame in the car video. The car license plate boxed in dashes is the selected ROI.
sequence is shown in Fig. 6. Here the car license plate is selected as our ROI. We used five LR images with frame 1 (the referenced frame) shown in Fig. 7(a) to obtain a resolution enhancement factor of 2. The camera blur kernel is assumed to be a $5 \times 5$ Gaussian kernel with a standard deviation equal to 0.5. The selected threshold for Prewit operators is 60. Fig. 7(b–d) shows the bilinear interpolated image, the SR result using HMRF regularization, and the result with HWMRF regularization, respectively. Detailed regions cropped from Fig. 7(b–d) are

![Fig. 7](image) Results of super-resolution reconstruction using different regularizations. (a) Frame 1 of the surveillance sequence. (b) Bilinear interpolation. (c) SR result with HMRF regularization. (d) SR result with HWMRF regularization.

![Fig. 8](image) (a–c) Detailed regions cropped from Fig. 7(b–d), respectively.
shown in Fig. 8(a–c), respectively. Evidently, the proposed algorithm using HWMRF regularization has better visual quality than the conventional MAP algorithm with HMRF regularization.

Our second real experiment uses the disk surveillance sequence. Here we used five frames with frame 1 (the referenced frame) which is shown in Fig. 9(a). The camera blur kernel is assumed to be a $5 \times 5$ Gaussian kernel with

---

**Fig. 9.** Results of super-resolution reconstruction using different regularizations. (a) Frame 1 of the surveillance sequence. (b) Bilinear interpolation. (c) SR result with HMRF regularization. (d) SR result with HWMRF regularization. (e) SR result with HWMRF regularization of frame 3. (f) SR result with HWMRF regularization of frame 5.
a standard deviation equal to 0.5. The selected threshold for Prewit operators is set to 30. Fig. 9(b–d) shows the bilinear interpolated image, the SR result using HMRF regularization, and the result with HWMRF regularization, respectively. Detailed regions cropped from Fig. 9(b–d) are shown in Fig. 10(a–c), respectively. It is apparent that the proposed algorithm using HWMRF regularization obtained more desirable result than the bilinear interpolation and the conventional MAP algorithm with HMRF regularization. Other frames of the surveillance sequence can also be selected as the referenced frame, so that we can obtain a HR surveillance sequence. SR reconstruction results of frames 3 and 5 are showed in Fig. 9(e–f), respectively. Detailed regions cropped from Fig. 9(e–f) are shown in Fig. 10(d–e), respectively.

6. Conclusions

This paper proposes an edge-preserving MAP estimation based super-resolution algorithm by using the weighted Markov random field regularization for a ROI from several low-resolution surveillance images. The motion of the ROI during the sequence is assumed to
Acknowledgments

We want to thank the helpful comments and suggestions from the anonymous reviewers. This work was supported by Major State Basic Research Development Program (973 Program) of China under Grant No. 2009CB723905, the National High Technology Research and Development Program (863 Program) of China under Grant Nos.2009AA12Z114, 2007AA12Z148, the National Natural Science Foundation of China under Grant Nos. 40930532, 40801182, 40771139, 40523005, 40721001.

References