A Discontinuity Adaptive Method for Super-Resolution of License Plates

K.V. Suresh and A.N. Rajagopalan

Image Processing and Computer Vision Laboratory,
Department of Electrical Engineering,
Indian Institute of Technology Madras,
Chennai - 600 036, India
sureshkvsit@yahoo.com, raju@ee.iitm.ac.in

Abstract. In this paper, a super-resolution algorithm tailored to enhance license plate numbers of moving vehicles in real traffic videos is proposed. The algorithm uses the information available from multiple, sub-pixel shifted, and noisy low-resolution observations to reconstruct a high-resolution image of the number plate. The image to be super-resolved is modeled as a Markov random field and is estimated from the low-resolution observations by a graduated non-convexity optimization procedure. To preserve edges in the reconstructed number plate for better readability, a discontinuity adaptive regularizer is proposed. Experimental results are given on several real traffic sequences to demonstrate the edge preserving capability of the proposed method and its robustness to potential errors in motion and blur estimates. The method is computationally efficient as all operations are implemented locally in the image domain.

1 Introduction

Intelligent Transport Systems (ITS) that combine electronics, information, communication, and network technologies are being increasingly used to address traffic problems in developed as well as developing countries [1]. One of the important goals of ITS is to decipher the identity of vehicles to enable monitoring of offenses and crimes on public routes. If a low-resolution video surveillance system captures an untoward incident on the road, a post-facto analysis of the stored video may be required. However, due to image degradation, information about the identity of the vehicles involved in the incident may not be easily derivable. For improving the readability of license plate text, a method is suggested in [2] that enhances only the character pixels while de-emphasizing the background pixels. Cui et al. [3] have presented a multi-frame-based binarization scheme for the extraction and enhancement of characters in license plates. Sato et al. [4] present a sub-pixel interpolation-based video text enhancement scheme. But interpolation cannot restore the high frequency components lost during sampling.

The video quality degrades due to various reasons such as motion blur, distance to camera, and noise. Cost considerations also dictate the resolution of

P. Kalra and S. Peleg (Eds.): ICVGIP 2006, LNCS 4338, pp. 25–34, 2006.

surveillance cameras. Super-resolution is a process in which a high-resolution (HR) image is constructed from a set of sub-pixel shifted low-resolution (LR) images. Fundamentally, the task involves dealiasing and deblurring [5]. For the problem on hand, since there is relative motion between the camera and the vehicle, one can use sub-pixel motion information for enhancing the text in traffic videos. In [6], a Bayesian super-resolution algorithm based on the simultaneous autoregressive model developed for text image sequences is used to enhance license plates. In [7], a method for generating an HR slow-motion sequence from compressed video is suggested, in which an area of interest such as the license plate is slowed down and super-resolved. Miravet and Rodriguez [8] use neural networks to perform super-resolution of license plates. A learning-based framework has been proposed in [9] for zooming the digits in a license plate.

In this paper, our aim is to propose a super-resolution algorithm suitable for enhancing license plate text in real traffic videos. This is a challenging problem for several reasons. The distance of the camera to the vehicle is typically large rendering it difficult for even humans to decipher the text. The low-resolution images are quite noisy, and blurred. Motion and blur estimates derived from such degraded images will not be correct. It is well-known fact that the performance of super-resolution algorithms is good only when these parameters are known accurately. The high-resolution license plate image is modeled as a Markov Random Field (MRF) and a maximum *a posteriori* (MAP) estimate of the super-resolved image is obtained, given the low-quality observations. The purpose behind modeling by MRF which is a statistical characterization is to lend robustness to errors in motion and blur estimates during the reconstruction process. Since our objective is to improve readability of the license plate text, we propose a discontinuity adaptive MRF (DAMRF) prior in which the degree of interaction between pixels across edges is adjusted adaptively. Because this prior is non-convex, we use Graduated Non-Convexity (GNC) which is a deterministic annealing algorithm for performing optimization. All matrix operations are implemented as local image operations for computational speed-up. The performance of the proposed method is found to be quite good when tested on real traffic video sequences.

2 Problem Formulation

The relation between a lexicographically ordered low-resolution observation and the original high-resolution image can be expressed in matrix formulation as

$$\underline{y_r} = DH_rW_r\underline{x} + \underline{n_r}, \qquad 1 \leq r \leq m \tag{1}$$

Here, $\underline{x}$ is the original HR image of dimension N_1N_2 x 1, $\underline{y_r}$ is the r^{th} LR image of dimension M_1M_2 x 1, D is a down-sampling matrix of dimension M_1M_2 x N_1N_2. Matrix H_r is the camera defocus blur matrix, and W_r is the geometric warping matrix for the r^{th} frame. Each of these matrices is of dimension N_1N_2 x N_1N_2. The term $\underline{n_r}$ is the noise in the r^{th} frame. We assume that there are m number of LR observations i.e., $1 \leq r \leq m$.

Solving for $\underline{x}$ in Eq. (1) given the observations $\underline{y_r}$ is an ill-posed inverse problem. Because the blur operator may exhibit zeros in the frequency domain rendering the process non-invertible. At high frequencies, there will be excessive noise amplification since the transfer function of the blurring operator is low-pass in nature. Moreover, the presence of noise in the observation process can result in an observation sequence which is inconsistent with any scene. Hence, it is important to use *a priori* information about $\underline{x}$ that will reduce the space of solutions which conforms to the observed data. The Bayesian MAP formulation allows for incorporation of prior knowledge about $\underline{x}$ to improve robustness during the reconstruction process.

The MAP estimate of the super-resolved image $\underline{x}$ given m low-resolution images is given by

$$\hat{\underline{x}} = \arg\max_{\underline{x}} \{P(\underline{x}|\underline{y_1}, \cdots \underline{y_m})\} \tag{2}$$

Using Bayes' rule and taking the logarithm of the posterior probability, the MAP estimate of $\underline{x}$ is given by

$$\hat{\underline{x}} = \arg\max_{\underline{x}} \{\log[P(\underline{y_1}, \cdots \underline{y_m}|\underline{x})] + \log P(\underline{x})\} \tag{3}$$

We need to specify the prior image density $P(\underline{x})$ and the conditional density $P(\underline{y_1}, \cdots \underline{y_m}|\underline{x})$. Using the observation model in Eq. (1) and the fact that the noise fields are statistically independent of X and as well as each other, we have

$$P(\underline{y_1}, \cdots \underline{y_m}|\underline{x}) = \frac{1}{(2\pi\sigma^2)^{m\frac{M_1 M_2}{2}}} \exp\left\{-\sum_{r=1}^{m} \frac{||\underline{y_r} - DH_r W_r \underline{x}||^2}{2\sigma^2}\right\} \tag{4}$$

where σ^2 is the variance of the observation noise.

Using Eq. (4) in Eq. (3) and neglecting constant terms, the MAP estimate can be equivalently written as

$$\hat{\underline{x}} = \arg\min_{\underline{x}} \left\{\sum_{r=1}^{m} \frac{||\underline{y_r} - DH_r W_r \underline{x}||^2}{2\sigma^2} - \log P(\underline{x})\right\} \tag{5}$$

3 Discontinuity Adaptive MRF (DAMRF) Prior

We model the super-resolved image to be estimated as a Markov random field because it provides a foundation for the characterization of contextual constraints and the densities of the probability distributions of interacting features in images.

MRF theory helps in analyzing the spatial dependencies of physical phenomena. Let F be a random field over an N x N lattice of sites $L = (i, j) : 1 \leq i, j \leq N$. The random field F is said to be an MRF on L with respect to a neighborhood system η if

1. $P(F = f) > 0, \forall f \in \mathcal{F}$
2. $P[F_{i,j} = f_{i,j} | F_{k,l} = f_{k,l}, \forall (k,l) \neq (i,j)] = P[F_{i,j} = f_{i,j} | F_{k,l} = f_{k,l}, (k,l) \in \eta_{i,j}]$

where $\eta_{i,j}$ is the neighborhood of the site (i,j) and $\mathcal{F}$ denotes the configuration space. It is natural to expect that the image intensity at a pixel will not depend on the image data outside its neighborhood when the image data on its neighborhood are given. MRF image models even with first order neighborhood system are known to be powerful.

The practical use of MRF models can be largely ascribed to the equivalence between MRFs and Gibbs Random Field (GRF) established by Hammersely and Clifford [10]. The theorem states that F is an MRF on L with respect to neighborhood η if and only if F is a Gibbs random field on L. i.e.,

$$P[\underline{F} = \underline{f}] = \frac{1}{Z} \exp\{-U(\underline{f})\} \tag{6}$$

where Z is the partition function given by $Z = \sum_f \exp\{-U(\underline{f})\}$ and $U(\underline{f})$ is the energy function which is given by

$$U(\underline{f}) = \sum_{c \in C} V_c(\underline{f}) \tag{7}$$

Here, c is called the clique of the pair (L, η) which is a subset of sites in L in which all pairs of sites are mutual neighbors. The set C is the set of all cliques. Since we model the HR image X as an MRF, we can write

$$P[\underline{X} = \underline{x}] = \frac{1}{Z} \exp\{-U(\underline{x})\} \tag{8}$$

where

$$U(\underline{x}) = \sum_{c \in C} V_c(\underline{x}) \tag{9}$$

The choice of the clique potential $V_c(\underline{x})$ is crucial as it embeds important prior information about the image to be reconstructed. The prior model can be chosen as

$$\sum_{c \in C} V_c(\underline{x}) = \sum_{c \in C} g(d_c \underline{x})$$

where $d_c \underline{x}$ is a local spatial activity measure of the image and has a small value in smooth regions and a large value at edges. A common choice for the prior model is a Gauss-Markov random field model [11] which has the form

$$g(n) = n^2$$

However, this image model can result in a blurred estimate of the super-resolved license plate, particularly along edges due to over-smoothing. Geman and Geman [10] introduced the concept of line fields which helps in preserving edges. But

the use of line fields makes the energy function non-differentiable. Schultz et al. [12] have used a discontinuity preserving model of the form

$$g(n) = \begin{cases} n^2, & |n| \leq T \\ 2T(|n| - T) + T^2, & |n| > T \end{cases}$$

where T is the threshold parameter separating the quadratic and linear regions. The threshold which is dependent on factors like image content and noise has to be appropriately tuned for every new case. This threshold when fixed at low values lets in noise and at high values penalizes weak edges.

To improve the readability of the license plate, we propose to use a discontinuity adaptive MRF (DAMRF) model in which the degree of interaction between pixels across edges is adjusted adaptively in order to preserve discontinuities. A necessary condition for any regularization model to be adaptive to discontinuities [13] is

$$\lim_{n\to\infty} |g'(n)| = \lim_{n\to\infty} |2nh(n)| = C \tag{10}$$

where n is the difference between neighboring pixel values and $C \in [0, \infty)$ is a constant. We propose to choose $g(n)$ as

$$g(n) = \gamma - \gamma e^{-n^2/\gamma} \tag{11}$$

Fig. 1 shows the function defined by Eq. (11). It is convex in the band $B_\gamma = \left(-\sqrt{\gamma/2}, \sqrt{\gamma/2}\right)$ and non-convex outside. The DA function allows the smoothing strength to increase monotonically as n increases within the band B_γ thus smoothing out noise. Outside this band, smoothing decreases as n increases thereby preserving the discontinuities.

Using the DA prior function and assuming a first-order neighborhood for MRF, we can write

$$\begin{aligned} \sum_{c\in C} V_c(\underline{x}) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} 4 * \gamma - \gamma \exp\{-[x(i,j) - x(i,j-1)]^2/\gamma\} \\ -\gamma \exp\{-[x(i,j) - x(i,j+1)]^2/\gamma\} - \gamma \exp\{-[x(i,j) - x(i-1,j)]^2/\gamma\} \\ -\gamma \exp\{-[x(i,j) - x(i+1,j)]^2/\gamma\} \end{aligned} \tag{12}$$

Using Eqs. (12), (9), and (8) in Eq. (5) and finding the gradient at the n^{th} iteration, we get

$$\text{grad}^{(n)} = \frac{1}{\sigma^2} \sum_{r=1}^{m} W_r^T H_r{}^T D^T (D H_r W_r \underline{x} - \underline{y_r}) + \lambda G^{(n)} \tag{13}$$

where λ is the regularization parameter and the gradient at (k, l) is given by

$$\begin{aligned} G^{(n)}(k,l) = 2[x(i,j) - x(i,j-1)] \exp\{-[x(i,j) - x(i,j-1)]^2/\gamma\} + \\ 2[x(i,j) - x(i,j+1)] \exp\{-[x(i,j) - x(i,j+1)]^2/\gamma\} + \\ 2[x(i,j) - x(i-1,j)] \exp\{-[x(i,j) - x(i-1,j)]^2/\gamma\} + \\ 2[x(i,j) - x(i+1,j)] \exp\{-[x(i,j) - x(i+1,j)]^2/\gamma\} \end{aligned} \tag{14}$$

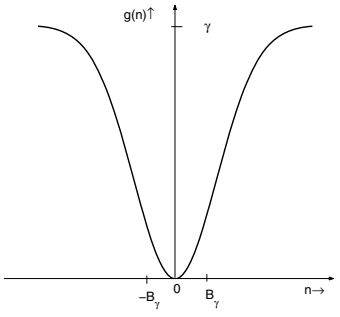


Fig. 1. A discontinuity adaptive function

4 Optimization Using Graduated Non-convexity

The DA function is non-convex and annealing can be used to overcome the problem of local minima. There are two types of annealing: deterministic and stochastic. We use a deterministic annealing method called Graduated Non-Convexity (GNC) algorithm for optimization [13]. The idea of GNC is to start with a strictly convex cost function by choosing a large value for γ and to find a unique minimum using gradient descent in the first phase. This value is then used as the initial value for the next phase of minimization with a smaller γ. These steps are repeated by lowering the value of γ until convergence. It finds a good solution with much less computational cost.

Algorithm. Super-resolution using GNC

Require: Observations $\{Y_i\}$, blur kernels, and motion parameters.

1: Calculate $X^{(0)}$ as the average of the bilinearly up-sampled and aligned images.

2: Choose a convex $\gamma^{(0)} = 2v$, where v is the maximum value of the gradient along the x and y directions in the initial estimate $X^{(0)}$.

3: $n = 0$

4: Do
   a. Update $X^{(n)}$ using $X^{(n+1)} = X^{(n)} - \alpha\,\mathrm{grad}^{(n)}$
   b. Set $n = n + 1$;
   c. If $(\mathrm{norm}(X^{(n)} - X^{(n-1)}) < \epsilon)$ set $\gamma^{(n)}$=max $[\gamma_{\mathrm{target}}, k\gamma^{(n-1)}]$;
   UNTIL $(\mathrm{norm}(X^{(n)} - X^{(n-1)}) < \epsilon)$ and $(\gamma^{(n)} = \gamma_{\mathrm{target}})$;

5: Set $\hat{X} = X^{(n)}$
   where α is the step size, ϵ is a constant for testing convergence, and k is a factor that takes $\gamma^{(n)}$ slowly towards γ_{target}.

Calculation of the gradient in Eq. (13) involves operations on large matrices which can be computationally very intensive. The matrices W_r, H_r, and D, and their transposes are implemented using only simple local image operations as follows thereby yielding a considerable speed-up.

- D is implemented by averaging q^2 pixels in the higher dimension to calculate each pixel in the lower dimension where q is the resolution factor.
- H_r is implemented by convolving the image with the respective blur kernels.
- W_r is implemented by warping the image using bilinear interpolation.
- D^T is implemented by spreading equally the intensity value in the lower dimension to the q^2 pixels in the higher dimension.
- H_r^T is implemented by convolution of the image with the flipped kernel. i.e., if $h(i,j)$ is the imaging blur kernel, then the flipped kernel $\hat{h}$ satisfies $\hat{h}(i,j) = h(-i,-j), \quad \forall (i,j)$.
- W_r^T is implemented by backward warping if W_r is implemented by forward warping.

Note that for implementation of matrix D in image domain, we need (q^2-1) additions and one multiplication (by $\frac{1}{q^2}$) whereas D^T needs one multiplication (by $\frac{1}{q^2}$) to calculate each pixel. The warping operation is typically performed using bilinear interpolation. Each pixel value in the warped image is calculated from its four neighboring pixels using the interpolation coefficients. Hence to implement matrices W_r and W_r^T in image domain, we need 7 additions and 8 multiplications (except at the borders) to determine each pixel. The number of computations for blurring an image depends on the size of the blur kernel. If we denote the kernel size as bl_size, then we need bl_size^2 multiplications and $(bl\_size^2 - 1)$ additions to compute each pixel.

The overall computational advantage that can be derived by implementing the proposed algorithm using local image domain operations instead of large matrix multiplications is given in Table 1. The table gives comparisons for implementation of D, H_r, W_r, and their transposes. We assume the dimension of the HR image to be $N \times N$ and that of the LR image to be $M \times M$. The blur kernel size is denoted by bl_size and q is the resolution factor. Note that, there is a substantial gain in implementing using local image operations.

Table 1. Computations required for W_r, H_r, D, and their transposes

Operation	**Matrix domain computations**	**Image domain computations**
W_r, W_r^T	$N^2 \times N^2$ multiplications $N^2 \times (N^2-1)$ additions	$N^2 \times 8$ multiplications $N^2 \times 7$ additions
H_r, H_r^T	$N^2 \times N^2$ multiplications $N^2 \times (N^2-1)$ additions	$N^2 \times bl\_size^2$ multiplications $N^2 \times (bl\_size^2-1)$ additions
D	$M^2 \times N^2$ multiplications $M^2 \times (N^2-1)$ additions	$M^2 \times 1$ multiplications $M^2 \times (q^2-1)$ additions
D^T	$N^2 \times M^2$ multiplications $N^2 \times (M^2-1)$ additions	$M^2 \times 1$ multiplications -Nil-

5 Experimental Results

In this section, we demonstrate the performance of the proposed method for super-resolving license plates and also compare it with other techniques. In our

experiments, we considered resolution improvement by a factor of 2. The values chosen for the various parameters were $\lambda = 0.005$, $\gamma^{(0)} = 300$, $\gamma_{\text{target}} = 10$, $k = 0.95$, and $\alpha = 6$. We considered real data for testing our method. For this purpose, video frames (25 frames/second) of a busy traffic way were captured using a SONY handycam. The data was gathered from a flyover of height about 20 feet. The viewing angle of the camera relative to the ground was about 45^o. The movement of vehicles was away from the camera. Because we use successive frames, scaling is negligible and is ignored. Since the vehicles were moving away from the camera and roughly along a straight line there was no rotation. Our objective is to go beyond the resolution of the camera to enhance the license plate region by using the motion information in the captured observations.

In the first example, the license plate (of size 16×58 pixels) of a moving car was cropped from four consecutive frames of the traffic video and these low-quality frames are shown in Figs. 2(a)-2(d). The sub-pixel motion corresponding to LR frames was computed using [14]. The resultant motion estimates were fed as input to different super-resolution techniques, namely the LS method [15], the GMRF method [11], the HMRF method [12], and the proposed method. Note that these motion estimates are not accurate since they are computed from noisy, and aliased observations. The assumption of Gaussian PSF for the camera defocus blur is also an approximation. Results corresponding to each of the above methods is shown in Fig. 2. The reconstructed image using the LS technique (Fig. 2(e)) is poor as it is very sensitive to errors in motion and blur estimates. The output of the GMRF algorithm (Fig. 2(f)) is quite blurred and some of the numbers are not at all discernible. HMRF performs relatively better (Fig. 2(g)) but some of the numbers are not easily readable. For example, the second digit '3' can be confused with '9' while the last digit '4' can be misinterpreted as '6'. In comparison, the proposed DAMRF algorithm yields the best result with distinctly defined edges as shown in Fig. 2(h). The license plate number (K 8354) can be read clearly without any ambiguity.

In the next example, the license plate of another car was cropped from four consecutive frames (Figs. 3(a)-3(d)). Note that the visual quality of these plates is very poor. The output corresponding to different super-resolution methods is given in Figs. 3(e)-3(h). We again observe that the reconstructed image using DAMRF is significantly better compared to existing methods. The text on the

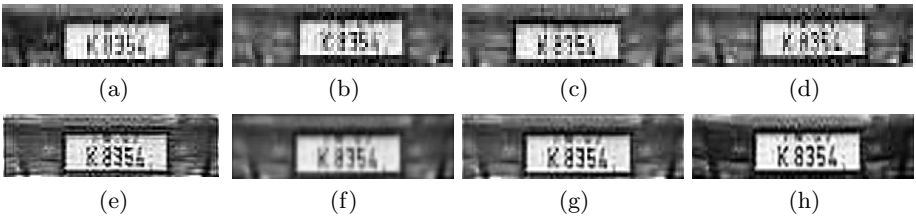
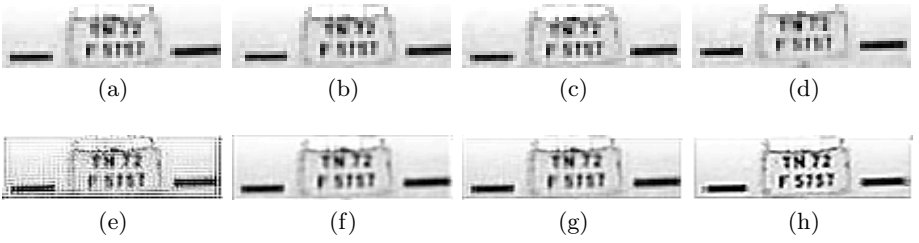


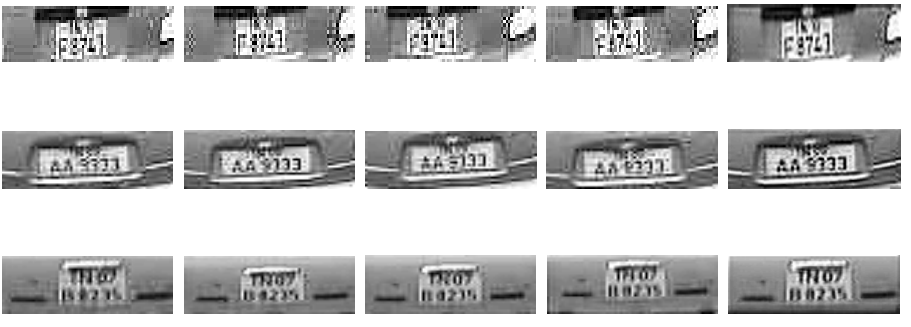
Fig. 2. (a)-(d) Cropped license plates. Super-resolved image using (e) LS, (f) GMRF, (g) HMRF, and (h) DAMRF.



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 3. (a)-(d) Cropped license plates. Super-resolved image using (e) LS, (f) GMRF, (g) HMRF, and (h) DAMRF.



(a) (b) (c) (d) (e)

Fig. 4. (a)–(d) Low resolution observations. (e) Super-resolved images using the proposed method.

license plate comes out clearly in the super-resolved image using the proposed method.

In Fig. 4 we have given results corresponding to the license plates of some more cars. Note that in all the cases, the readability of the number plate improves significantly after performing super-resolution on the captured video frames using the proposed method.

6 Conclusions

A robust super-resolution algorithm using a discontinuity adaptive prior is proposed to enhance the license plate text of moving vehicles. The algorithm fuses the information available from multiple observations of a vehicle to obtain a high quality license plate image. The high-resolution image is modeled as an MRF and is estimated using graduated non-convexity. The effectiveness of the proposed method was demonstrated on many real traffic video sequences. The proposed DAMRF method is robust to errors in motion and blur estimates and preserves the edges in the reconstructed license plate text.

References

1. Rajagopalan, A.N., Chellappa, R.: Vehicle detection and tracking in video. In: Proc. of Intl. Conf. on Image Process. **1** (2000) 351–354
2. Zhang, Y., Zhang, C.: A new algorithm for character segmentation of license Plate. In: Proc. of IEEE Intelligent Vehicles Symp. (2003) 106–109
3. Cui, Y., Huang, Q.: Character extraction of license plates from video. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition. (1997) 502–507
4. Sato, T., Kanade, T., Hughes, E., Smith, M., Satoh, S.: Video OCR: Indexing digital news libraries by recognition of superimposed caption. ACM Multimedia Systems Special Issue on Video Libraries. **7** (1999) 385–395
5. Chaudhuri, S.: Super-resolution imaging. Kluwer Academic, USA. (2001)
6. Cortijo, F.J., Villena, S., Molina, R., Katsaggelos, A.: Bayesian super-resolution of text image sequences from low-resolution observations. In: IEEE Intl. Symp. on Signal Process. and its Application. (2003) 421–424
7. Chaudhuri, S., Taur, D.R.: High-resolution slow-motion sequencing - How to generate a slow-motion sequence from a bit stream. IEEE Signal Process. Mag. **22** (2005) 16-24
8. Miravet, C., Rodriguez, F.B.: A hybrid MLP-PNN architecture for fast image super-resolution. In: Intl. Conf. on Neural Information Process. (2003) 417–424
9. Rajaram, S., Gupta, M.D., Petrovic, N., Huang, T.S.: Learning-based nonparametric image super-resolution. EURASIP Journal on Applied Signal Process. (2006)
10. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Trans. on Pattern Anal. and Mach. Intell. **6** (1984) 721–741
11. Hardie, R.C., Barnard, K., Armstrong, E.E.: Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. IEEE Trans. on Image Process. **6** (1997) 1621–1632
12. Schultz, R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. on Image Process. **5** (1996) 996-1011
13. Li, S.Z.: Markov random field modeling in computer vision. Springer-Verlag, Tokyo. (1995)
14. Irani, M., Peleg, S.: Improving resolution by image registration. CVGIP: Graph. Models and Image Process. **53** (1991) 231–239
15. Zomet, A., Peleg, S.: Super-resolution from multiple images having arbitrary mutual motion. Super-resolution Imaging. ed., Chaudhuri, S., Kluwer Academic. (2001) 195–209