Superresolution from Occluded Scenes

Wataru Fukuda, Atsunori Kanemura, Shin-ich Maeda, and Shin Ishii

Graduate School of Informatics, Kyoto University, Kyoto 611-0011, Japan
{fukuda,atsu-kan,ichi,ishii}@sys.i.kyoto-u.ac.jp

Abstract. We propose a Bayesian image superresolution method that estimates a high-resolution background image from a sequence of occluded observations. We assume that the occlusions have spatial and temporal continuities. Such assumptions would be plausible, for example, when satellite images are occluded by clouds or when a tourist site is obstructed by people. Although the exact inference of our model is difficult, an efficient superresolution algorithm is derived by using a variational Bayes technique. Experiments show that our superresolution method performs better than existing methods that do not assume the occlusions or that assume the occlusions but do not assume the temporal continuities of the occlusions.

Keywords: Occlusion removal, image superresolution, variational Bayes.

1 Introduction

Superresolution is an image processing method that estimates a high-resolution (HR) image from a sequence of low-resolution (LR) images of a single scene [1,2,3,4]. In this study, we focus on a superresolution method that deals with occluded observations. Occlusions are typical, for example, in satellite images, where clouds can obstruct the land surface, or in pictures taken at tourist sites, where people may disturb taking a clean image.

There are two strategies to superresolving an HR image from given occluded LR images. One way is to identify the occlusions and use the unoccluded regions of the observations for superresolution [1,5] and the other is to employ a robust cost function that automatically discards inconsistent parts of the observations [2]. The former, which we adopt in this study, has several advantages over the latter; it can incorporate explicit prior knowledge about the occlusion and clean HR image, and desirable properties on them can be incorporated. Furthermore, by representing the existence of occlusions in a probabilistic manner, it can produce better estimates of the clean HR image by incorporating the uncertainty whether a pixel in an observed image is occluded or not.

The estimation of a clean HR image and the detection of occlusions are closely related; if we know one, we are able to accurately estimate the other. Consequently, a separated treatment of the estimation of the clean HR image and the identification of the obstructers (e.g. as in [6]) would prevent us from accurate estimation. Therefore, we conduct their simultaneous estimation based

C.S. Leung, M. Lee, and J.H. Chan (Eds.): ICONIP 2009, Part II, LNCS 5864, pp. 19–27, 2009.

on a consistent probabilistic model incorporating prior knowledge about the HR image and occlusions; in other words, we estimate posterior distributions of the clean HR image and obstructers conditioned on given observations based on a joint distribution of the clean HR image, the obstructers, and the given observations. However, exact Bayesian inference in our model is difficult due to the high dimensionality of the variables; thus we employ variational Bayes and Laplace approximation to obtain a practical algorithm.

2 Problem Formulation

Suppose we have T LR images $\mathbf{y} = \{\mathbf{y}^{(1)}, \cdots, \mathbf{y}^{(T)}\}$ that are consecutive and partly occluded observations of the same scene, where $\mathbf{y}^{(t)}$ is a vector of length P_{L} obtained by raster-scanning of the image. Our aim is to reconstruct an HR image $\mathbf{x}$, which is a vector of length P_{H} ($> P_{\mathrm{L}}$) from its observations $\mathbf{y}$.

The observed images are assumed to be generated from the HR image by i) applying shift and rotation motions, ii) convolving with a point spread function (PSF), iii) downscaling, and iv) adding Gaussian noise. To represent possible occlusions, we further assume the variance of the noise process is space-*variant* depending on the existence of occluders [1]. Given the occluders and $\mathbf{x}$, $\mathbf{y}^{(t)}$ is assumed to be generated by the following linear equation:

$$\mathbf{y}^{(t)} = W^{(t)}\mathbf{x} + \boldsymbol{\epsilon}^{(t)} \quad \text{for } t = 1, \ldots, T, \tag{1}$$

where $W^{(t)}$ denotes a $P_{\mathrm{L}} \times P_{\mathrm{H}}$ observation transformation matrix that reflects the assumptions i)–iii), and the noise term $\boldsymbol{\epsilon}^{(t)}$ obeys *non*-iid Gaussian distribution whose variance is space-variant according to the occlusion pattern. In this study, we assume for simplicity the matrices $W^{(t)}$ are known.

3 Hierarchical Bayesian Model

According to a Bayesian treatment of superresolution, the HR image is estimated by the mean of its posterior distribution $p(\mathbf{x}|\mathbf{y})$,

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\,\mathrm{d}\mathbf{x}}. \tag{2}$$

We define a hierarchical model by introducing hidden variables $\mathbf{z} = \{\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(T)}\}$ that represent occlusions on the LR images. Each $\mathbf{z}^{(t)} \in \{-1, +1\}^{P_{\mathrm{L}}}$ has the same length as the corresponding LR image $\mathbf{y}^{(t)}$, and $z_i^{(t)} = -1$ denotes that the ith pixel of $\mathbf{y}^{(t)}$ is occluded, whereas $z_i^{(t)} = +1$ denotes the corresponding pixel is intact. Therefore, the two-dimensional configuration of $\mathbf{z}^{(t)}$ shows the shape of the occlusion on the tth LR image. We further introduce hidden variables $\boldsymbol{\theta}$ which represent the temporal movements of the occluder. In the following subsections we define probability distributions that constitute our model (2).

3.1 Observation Model

The observation process is probabilistically given by

$$p(\mathbf{y}|\mathbf{x},\mathbf{z}) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{y}^{(t)}|W^{(t)}\mathbf{x}, B(\mathbf{z}^{(t)})^{-1}), \tag{3}$$

where $B(\mathbf{z}^{(t)})$ is a $P_{\mathrm{L}} \times P_{\mathrm{L}}$ diagonal precision matrix whose elements $\beta_i(z_i^{(t)})$ $(i = 1, \ldots, P_{\mathrm{L}})$ are defined by

$$\beta_i(z) = \begin{cases} \beta_{\mathrm{H}}, & z = +1, \\ \beta_{\mathrm{L}}, & z = -1. \end{cases} \tag{4}$$

We assume $\beta_{\mathrm{H}} > \beta_{\mathrm{L}}$ so that $z_i^{(t)} = +1$ indicates high reliability (unoccluded) on the ith pixel of the tth observation whereas $z_i^{(t)} = -1$ low reliability (occluded). This space-variant variance model is the same as [1].

3.2 Image Prior

We assume a Gaussian image prior for the HR image

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mathbf{0}, (\rho A)^{-1}), \tag{5}$$

where the matrix A is a first-order difference operator and ρ is a scalar that controls the strength of smoothness. This prior represents our *a priori* knowledge that the HR image would be spatially smooth.

3.3 Prior for Occlusion Patterns

In order for a prior to represent the uniform motion of the occlusions, we assume a Markov property of the motion between the consecutive observations and introduce a two-dimensional vector $\boldsymbol{\theta}^{(t)} = [\theta_1^{(t)}, \theta_2^{(t)}]^{\top}$ that denotes the movement of the occlusion patterns between the tth and $(t+1)$th frames. Since we have assumed that each occluded region has a spatial continuity and moves smoothly across consecutive frames, we define

$$p(\mathbf{z}|\boldsymbol{\theta}) = p(\mathbf{z}^{(1)}) \prod_{t=1}^{T-1} p(\mathbf{z}^{(t+1)}|\mathbf{z}^{(t)}, \boldsymbol{\theta}^{(t)}) \tag{6}$$

$$= p(\mathbf{z}^{(1)}) \prod_{t=1}^{T-1} \frac{1}{Z_t} \exp\{-E(\mathbf{z}^{(t+1)}, \mathbf{z}^{(t)}, \boldsymbol{\theta}^{(t)})\}, \tag{7}$$

where Z_t is a normalizing constant and the energy function E is defined by

$$E(\mathbf{z}^{(t+1)}, \mathbf{z}^{(t)}, \boldsymbol{\theta}^{(t)})$$
$$= -J_{\mathrm{self}} \sum_i z_i^{(t+1)} - J_{\mathrm{inner}} \sum_{i \sim j} z_i^{(t+1)} z_j^{(t+1)} - J_{\mathrm{move}} \mathbf{z}^{(t+1)\top} G(\boldsymbol{\theta}^{(t)}) \mathbf{z}^{(t)}, \tag{8}$$

where $i \sim j$ means "i and j are adjacent pixels." The transition matrix $G(\boldsymbol{\theta}^{(t)})$ is a function of $\boldsymbol{\theta}^{(t)}$ and $G(\boldsymbol{\theta}^{(t)})\mathbf{z}^{(t)}$ represents the predicted occlusion pattern in the $(t+1)$th observation based on the occlusion pattern in the previous frame and the transition amount $\boldsymbol{\theta}^{(t)}$. The constants J_{self}, $J_{\text{inner}}(>0)$, and J_{move} (>0) represent the tendency of occlusion, a degree of correlation of neighbor pixels in the occlusion pattern, and a degree of correlation between $\mathbf{z}^{(t)}$ and $\mathbf{z}^{(t-1)}$ that determines how close the occlusion pattern $\mathbf{z}^{(t+1)}$ should be to the previous one, respectively. Although it is possible to estimate the parameters J_{self}, J_{inner} , and J_{move} from the given LR images, in this study they are given manually for focusing on the estimation of HR images.

The initial distribution $p(\mathbf{z}^{(1)})$ is defined as the same as $p(\mathbf{z}^{(t+1)}|\mathbf{z}^{(t)}, \boldsymbol{\theta}^{(t)})$ except the energy function contains only the fist two terms of (8).

3.4 Prior for Occlusion Transition

The occlusion transition parameters $\boldsymbol{\theta}$ are regarded as hidden variables that have a Markov property:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\theta}^{(1)}) \prod_{t=1}^{T-1} \mathcal{N}(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}, (rI)^{-1}), \tag{9}$$

where r is a scalar precision and I is the identity matrix. This prior asserts that the occluder is under the inertial law and likely to move towards the same direction as in the previous frame.

4 Approximate Bayesian Learning

Although the mean of the posterior $p(\mathbf{x}|\mathbf{y})$ is necessary for superresolution, the exact evaluation is intractable because we need to integrate out all the hidden variables and this cannot be performed analytically. Then, we employ the variational Bayes (VB) method and the Laplace approximation in order to derive a practical algorithm.

In the VB method, the joint posterior distribution $p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{y})$ is approximated by another distribution $q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$, which is called a trial distribution. The trial distribution is determined so as to minimize the Kullback-Leibler divergence between the true posterior and the trial distribution

$$D_{\text{KL}}(q||p) \equiv \left\langle \ln \frac{q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})}{p(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}|\mathbf{y})} \right\rangle_{q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})}, \tag{10}$$

where the square brackets $\langle \cdot \rangle_{q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})}$ denote an expectation operator with respect to $q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta})$. The Kullback-Leibler divergence is always nonnegative and equal to zero only when $q = p$.

To make the optimization tractable, we impose a factorization assumption on the trial distribution: $q(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = q(\mathbf{x}) \prod_{t,i} q(z_i^{(t)}) q(\boldsymbol{\theta})$ and the trial distribution is optimized by a coordinate-descent procedure. The coordinate-descent

minimization with respect to the factorized trial distribution gives closed-from solutions $q^*(\mathbf{x})$, $q^*(z_i^{(t)})$, $q^*(\boldsymbol{\theta})$.

As we will see later, $q^*(\mathbf{x})$ and $q^*(z_i^{(t)})$ are given by a Gaussian distribution and a Bernoulli distribution, respectively, and the integrals with respect to them are tractable. However, the integral with respect to $q^*(\boldsymbol{\theta})$ still remains intractable because $\ln q^*(\boldsymbol{\theta})$ is a complicated function of $\boldsymbol{\theta}$. To overcome this difficulty, we utilize the Laplace approximation; i.e., $q^*(\boldsymbol{\theta})$ is approximated by a Gaussian distribution, whose mean $\bar{\boldsymbol{\theta}}$ gives the maximum of $q^*(\boldsymbol{\theta})$.

4.1 Optimal Trial Distribution

Due to the VB approximation, the optimal trial distribution for $\mathbf{x}$ is found to be a Gaussian distribution

$$q^*(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \Sigma_{\mathbf{x}}), \tag{11}$$

where $\boldsymbol{\mu}_{\mathbf{x}} = \Sigma_{\mathbf{x}}(\sum_t W^{(t)\top}\langle B(\mathbf{z}^{(t)})\rangle \mathbf{y}^{(t)})$ and $\Sigma_{\mathbf{x}} = (\rho A + \sum_t W^{(t)\top}\langle B(\mathbf{z}^{(t)})\rangle W^{(t)})^{-1}$.

The direct evaluation of the probability distribution $q^*(\mathbf{x})$ is practically hard due to the inversion of the huge matrix $\Sigma_{\mathbf{x}}$. However, we can skip the direct calculation of $\Sigma_{\mathbf{x}}$ because all we have to know is $\boldsymbol{\mu}_{\mathbf{x}}$ for the final output image. Noting that the inverse of $\Sigma_{\mathbf{x}}$ is a sparse matrix, $\boldsymbol{\mu}_{\mathbf{x}}$ is obtained by a conjugate gradient (CG) algorithm [7].

Due to the VB and Laplace approximation, the optimal trial distribution for $\boldsymbol{\theta}$ becomes a Gaussian

$$q^*(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\theta}|\boldsymbol{\mu}_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}), \tag{12}$$

where $\boldsymbol{\mu}_{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}$ and $\Sigma_{\boldsymbol{\theta}} = \langle \bar{H} \rangle^{-1}$. Here, $\langle \bar{H} \rangle$ is the Hessian of the $\langle \ln q^*(\bar{\boldsymbol{\theta}}) \rangle_{q(\mathbf{z})}$.

The optimal trial distribution for $z_i^{(t)}$ is given by the following Bernoulli distribution:

$$q^*(z_i^{(t)}) = \nu_{ti}^{\frac{1}{2}(1+z_i^{(t)})}(1-\nu_{ti})^{\frac{1}{2}(1-z_i^{(t)})}. \tag{13}$$

The parameter ν_{ti}, which is the probability of the occlusion absence $q(z_i^{(t)} = +1)$, is calculated by a logistic sigmoid function $\nu_{ti} = \mathrm{sigmoid}(-2\lambda_{ti})$, where

$$\begin{aligned}\lambda_{ti} = {} & J_{\mathrm{self}} + J_{\mathrm{inner}} \sum_{j \in N(i)} \langle z_j^{(t)} \rangle + J_{\mathrm{move}}[\bar{G}^{(t-1)}\langle \mathbf{z}^{(t-1)} \rangle + \bar{G}^{(t)\top}\langle \mathbf{z}^{(t+1)} \rangle]_i \\ & + \frac{1}{2} J_{\mathrm{move}} \sum_{j,l,k} [\bar{G}^{(t-1)}_{\theta_k \theta_l ij} \langle z_j^{(t-1)} \rangle \Sigma_{\boldsymbol{\theta} kl} + \bar{G}^{(t)}_{\theta_k \theta_l ji} \langle z_j^{(t+1)} \rangle \Sigma_{\boldsymbol{\theta} kl}] \\ & + \frac{1}{4} \left[\ln(\beta_{\mathrm{H}}/\beta_{\mathrm{L}}) - (\beta_{\mathrm{H}} - \beta_{\mathrm{L}}) \langle e_{ti}^2 \rangle \right]. \end{aligned} \tag{14}$$

Here $N(i)$ is the set of the neighboring pixels of pixel i and partial derivatives are denoted by $\bar{G}^{(t)}_{\theta_k} = \partial G(\bar{\boldsymbol{\theta}}^{(t)})/\partial \theta_k^{(t)}$ and $\bar{G}^{(t)}_{\theta_l \theta_k} = \partial^2 G(\bar{\boldsymbol{\theta}}^{(t)})/\partial \theta_l^{(t)} \theta_k^{(t)}$. The first term

stems from the self connection, the second term from the inner-image connection, and the third and fourth terms from the occlusion movement. The last term is determined by the observations such that the baseline $(\beta_{\mathrm{H}} - \beta_{\mathrm{L}})^{-1} \ln(\beta_{\mathrm{H}}/\beta_{\mathrm{L}})$ is compared with the following expected squared error at the ith pixel of the tth observed image: $\langle e_{ti}^2 \rangle = (y_i^{(t)} - \mathbf{w}_i^{(t)\top} \boldsymbol{\mu}_{\mathbf{x}})^2 + \mathbf{w}_i^{(t)\top} \Sigma_{\mathbf{x}} \mathbf{w}_i^{(t)}$, where $\mathbf{w}_i^{(t)}$ is the ith row of $W^{(t)}$. Here, the first term represents the reconstruction error on the ith pixel of the tth image, and the second term represents the degree of uncertainty. This equation means that, if the reconstruction error at ith pixel is large, then the ith pixel is considered to be occluded.

The VB algorithm iteratively updates each component of the trial posterior, $q^*(\mathbf{x})$, $q^*(\mathbf{z})$ and $q^*(\boldsymbol{\theta})$, one by one by fixing the other components, until convergence. The algorithm is terminated when the relative change of $\boldsymbol{\mu}_{\mathbf{x}}$'s norm is smaller than a predetermined threshold 10^{-4}.

5 Experiments

5.1 Synthetic Dataset

The proposed algorithm was tested with synthetic data generated by the following procedure. A given original HR image $\mathbf{x}$ (Lena) was first transformed by translational and rotational motions, where the amounts of shift and rotation were randomly drawn from the respective uniform distributions $\mathrm{Unif}(-2, 2)$ and $\mathrm{Unif}(-4\pi/180, 4\pi/180)$, blurred with a Gaussian PSF with standard deviation 2, then downscaled by a factor of 4 in each direction, and corrupted by Gaussian noise whose variance is determined by the occlusion pattern $\mathbf{z}^{(t)}$. There were fifteen LR images. The occlusion pattern $\mathbf{z}^{(t)}$ is shown in Fig. 1. Noise SNR was set to 10 dB for the pixels where $z_i^{(t)} = -1$ (occluded pixels), while it was set to 40 dB for the pixels where $z_i^{(t)} = +1$ (intact pixels). The occlusion pattern on the first frame was generated by Gibbs sampling from the prior distribution $p(\mathbf{z}^{(1)})$, and the occlusion patterns on the other frames were generated according to $G(\boldsymbol{\theta}^{(t)}$. The values of $\boldsymbol{\theta}$ were drawn from a Gaussian distribution with a uniform mean vector. We fixed the model parameters at hand-tuned values: $\rho = 300$, $\gamma = 0.5$, $\beta_{\mathrm{H}} = 3 \times 10^5$, $\beta_{\mathrm{L}} = 10$.

Fig. 1. The binary occlusion patterns (black: large noise of 10 dB, white: small noise of 40 dB). The occlusion patterns were shifted according to $\boldsymbol{\theta}$.

(a) True (b) Without move (c) Proposed

Fig. 2. True and estimated occlusion patterns

(a) Observation (b) Proposed (c) Without move (d) Uniform

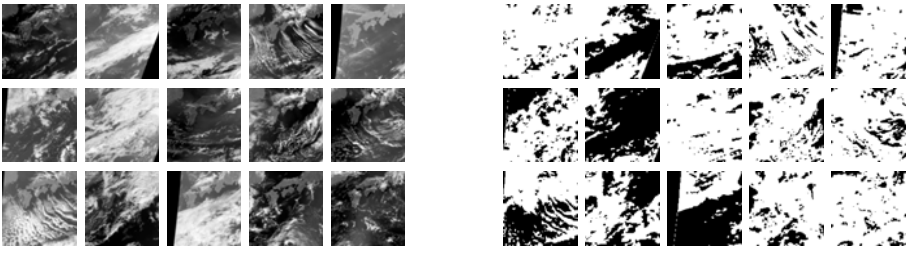
Fig. 3. Comparison of estimation results when the series of observed images suffer form the occlusions shown in Fig. 1. (a) One of the fifteen LR images. (b) Proposed (30.69 dB). (c) Without movement (29.13 dB). (d) Assumed uniform noise (28.66 dB).

Fig. 2 shows the true occlusion pattern and its estimates with/without assuming the movement of the occlusion patterns. The estimation results and close-up views of the region around the right eye of Lena are shown in Fig. 3. Fig. 3 shows (a) the first frame of the fifteen LR images, (b) the superresolved image estimated by our method, (c) the estimate without considering the temporal transition of the occlusions, (d) the estimate without assuming occlusions at all. Performance of the algorithm was measured by the peak signal-to-noise ratio (PSNR). As can be seen from the figure, our proposed method achieved the best PSNR (30.69 dB).

5.2 Cloud-Contaminated Dataset

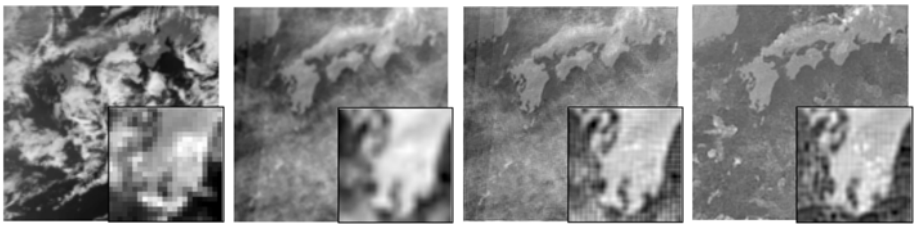
The second experiment is to obtain an HR image from a cloud-contaminated NOAA/AVHRR dataset published online by JADAS. We generated a set of LR images synthetically downsampled (by linear scaling of 4 to 1), again using the model with translational and rotational motions and blurred by Gaussinan PSF. The set of the LR images is shown in Fig. 4(a). The model parameters are assumed to be known.

Since the observations are not continuous in time, we do not assume occlusion transition here; the parameter for the occlusion prior J_{move} was set to 0. We fixed the model parameters at hand-tuned values: $J_{\mathrm{self}} = 0.02, J_{\mathrm{inner}} = 10, \rho = 400, \beta_{\mathrm{H}} = 8 \times 10^3$, and $\beta_{\mathrm{L}} = 10$.



(a) Observations (b) Estimated occlusion pattern

Fig. 4. Observed images and estimated occlusion patterns



(a) Observation (b) Median (c) Uniform (d) Proposed

Fig. 5. Estimation results using the cloud-contaminated images

Fig. 4(b) shows the estimated occlusion patterns. Fig. 5(a) presents an LR image, and Fig. 5(b) shows the observation median. Fig. 5(c) is the superresolved HR image by the simple method without assuming any occlusions. Fig. 5(d), we shows the result by the proposed algorithm. In this experiment, we cannot evaluate PSNR because we do not know the clean HR image. However, it appears that our model gives the best estimation in which the occlusions are removed successfully, and moreover, the quality of the HR image is enhanced. It is expected that the performance will be further improved if sequential observations are used.

6 Conclusion

We have proposed an image superresolution method for observations that are contaminated by occlusions. The key factor of our model is the hierarchical likelihood in which the hidden variables are utilized so as to represent the knowledge on the observation process. Uncertainty regarding the HR image, occlusion, and amounts of occlusion movement is effectively considered via Bayesian inference, and our model enables successful reconstruction by reducing the effects of occlusions and accomplishes the improvements in PSNR and the reconstructed HR image.

References

1. Kanemura, A., Maeda, S., Ishii, S.: Image superresolution under spatially structured noise. In: IEEE Int. Symp. Signal Processing and Information Technology, December 2007, pp. 275–280 (2007)
2. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Fast and robust multi-frame super-resolution. IEEE Trans. Image Process 13(10), 1327–1344 (2004)
3. Babacan, D., Molina, R., Katsaggelos, A.: Total variation super resolution using a variational approach. In: IEEE Int. Conf. Image Process, October 2008, pp. 641–644 (2008)
4. Pickup, L.C., Capel, D.P., Roberts, S.J., Zisserman, A.: Bayesian image super-resolution, continued. In: Advances in NIPS. MIT Press, Cambridge (2006)
5. Ivanovski, Z.A., Panovski, L., Karam, L.J.: Robust super-resolution based on pixel-level selectivity. In: Visual Commun. Image Process. Proc. SPIE, vol. 6077, pp. 80–87 (2006)
6. Melgani, F.: Contextual reconstruction of cloud-contaminated multitemporal multispectral images. IEEE Trans. Geosci. Remote Sens 44(2), 442–455 (2006)
7. Golub, G.H., van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)