IMAGE FUSION: A POWERFUL TOOL FOR OBJECT IDENTIFICATION

Filip Šroubek (sroubekf@utia.cas.cz), Jan Flusser (flusser@utia.cas.cz)
and Barbara Zitová (zitova@utia.cas.cz)
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, Praha 8, 182 08, Czech Republic

Abstract. Due to imperfections of imaging devices (optical degradations, limited resolution of CCD sensors) and instability of the observed scene (object motion, media turbulence), acquired images are often blurred, noisy and may exhibit insufficient spatial and/or temporal resolution. Such images are not suitable for object detection and recognition. Reliable detection requires recovering the original image. If multiple images of the scene are available, this can be achieved by image fusion.

In this chapter we review the respective methods of image fusion. We address all three major steps - image registration, blind deconvolution and resolution enhancement. Image registration brings the acquired images into spatial alignment, multiframe deconvolution estimates and removes the blur, and the spatial resolution of the image is increased by so-called superresolution fusion. Superresolution is the main topic of the chapter. We propose a unifying system that simultaneously estimates blurs and recovers the original undistorted image, all in high resolution, without any prior knowledge of the blurs and original image. We accomplish this by formulating the problem as constrained least squares energy minimization with appropriate regularization terms, which guarantees a close-to-perfect solution.

We demonstrate the performance of the method on many examples, namely on car license plate recognition and face recognition. Both of these tasks are of great importance in security and surveillance systems.

Key words: Image fusion, Multichannel systems, Blind deconvolution, Superresolution, Regularized energy minimization

1. Introduction

Imaging devices have limited achievable resolution due to many theoretical and practical restrictions. An original scene with a continuous intensity function $o[x, y]$ warps at the camera lens because of the scene motion and/or change of the camera position. In addition, several external effects blur images: atmospheric turbulence, camera lens, relative camera-scene motion, etc. We will call these effects *volatile blurs* to emphasize their unpredictable and transitory behavior, yet we will assume that we can model them as convolution with an unknown point spread function

(PSF) $v[x,y]$. This is a reasonable assumption if the original scene is flat and perpendicular to the optical axis. Finally, the CCD discretizes the images and produces a digitized noisy image $g[i,j]$ (frame). We refer to $g[i,j]$ as a *low-resolution (LR) image*, since the spatial resolution is too low to capture all the details of the original scene. In conclusion, the acquisition model becomes

$$g[i,j] = D((v * o[W(n_1,n_2)])[x,y]) + n[i,j]\,, \tag{1}$$

where $n[i,j]$ is additive noise and W denotes geometric deformation (spatial warping) of the image. Geometric deformations are partly caused by the fact that the image is a 2-D projection of a 3-D world, and partly by lens distortions and/or motion of the sensor during the acquisition. $D(\cdot) = S(g * \cdot)$ is the *decimation operator* that models the function of the CCD sensors. It consists of convolution with the *sensor PSF* $g[i,j]$ followed by the *sampling operator* S, which we define as multiplication by a sum of delta functions placed on an evenly spaced grid. The above model for one single observation $g[i,j]$ is extremely ill-posed. Instead of taking a single image we can take K ($K > 1$) images of the original scene and, in this way, partially overcome the equivocation of the problem. Hence we write

$$g_k[i,j] = D((v_k * o[W_k(n_1,n_2)])[x,y]) + n_k[i,j]\,, \tag{2}$$

where $k = 1,\ldots,K$ and D remains the same in all the acquisitions. In the perspective of this multiframe model, the original scene $o[x,y]$ is a single input and the acquired LR images $g_k[i,j]$ are multiple outputs. The model is therefore called a single input multiple output (SIMO) formation model. To our knowledge, this is the most accurate, state-of-the-art model, as it takes all possible degradations into account.

Because of many unknown parameters of the model, it is hard to analyze (automatically or visually) the images g_k and to detect and recognize objects in them. A very powerful strategy is offered by *image fusion*.

The term fusion means in general an approach to extraction of information adopted in several domains. The goal of image fusion is to integrate complementary information from all frames into one new image containing information the quality of which cannot be achieved otherwise. Here, the term “better quality” means less blur and geometric distortion, less noise, and higher spatial resolution. We may expect that object detection and recognition will be easier and more reliable when performed on the fused image. Regardless of the particular fusion algorithm, it is unrealistic to assume that the fused image can recover the original scene $o[x,y]$ exactly. A reasonable goal of the fusion is a discrete version of $o[x,y]$ that has higher spatial resolution than the resolution of the LR images and that is free of the volatile blurs. In the sequel, we will refer to this fused image as a *high resolution (HR) image* $f[i,j]$.

Fusion of images acquired according to the model (2) is a three-stage process – it consists of image registration (spatial alignment), which should compensate

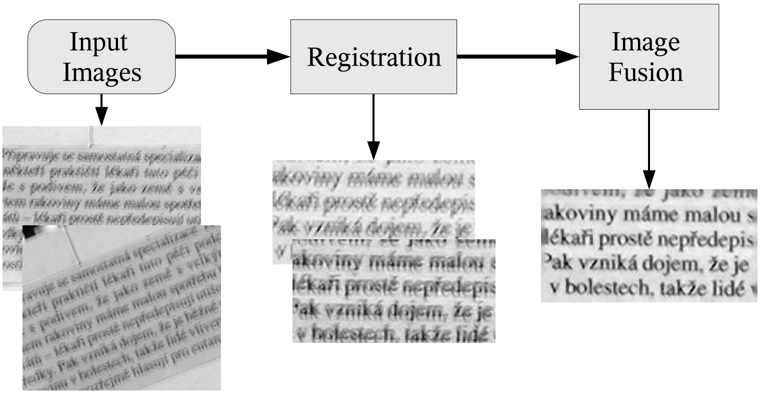


Figure 1. Image fusion in brief: Acquired images (left), registered frames (middle), fused image (right).

for geometric deformations W_k, followed by a *multichannel* (or multiframe) *blind deconvolution* (MBD) and *superresolution (SR) fusion*. The goal of MBD is to remove the impact of volatile blurs and the aim of SR is to increase spatial resolution of the fused image by a user-defined factor. While image registration is actually a separate procedure, we integrate both MBD and SR into a single step (see Fig 1), which we call *blind superresolution* (BSR). The approach presented in this chapter is one of the first attempts to solve BSR under realistic assumptions with only little *a priori* knowledge.

Image registration is a very important step of image fusion, because all MBD and SR methods require either perfectly aligned channels (which is not realistic) or allow at most small shift differences. Thus, the role of registration methods is to suppress large and complex geometric distortions. Image registration in general is a process of transforming two or more images into a geometrically equivalent form. From the mathematical point of view, it consists of approximating W_k^{-1} and of resampling the image. For images which are not blurred, registration has been extensively studied in the recent literature (see (Zitová and Flusser, 2003) for a survey). However, blurred images require special registration techniques. They can be, as well as the general-purpose registration methods, divided in two groups – global and landmark-based ones. Regardless of the particular technique, all feature extraction methods, similarity measures, and matching algorithms used in the registration process must be insensitive to image blurring.

Global methods do not search for particular landmarks in the images. They try to estimate directly the between-channel translation and rotation. In (Myles and Lobo, 1998) they proposed an iterative method which works well if a good initial estimate of the transformation parameters is available. In (Zhang et al., 2000; Zhang et al., 2002) the authors proposed to estimate the registration parameters by bringing the channels into canonical form. Since blur-invariant moments were used to define the normalization constraints, neither the type nor the level of the blur influences the parameter estimation. In (Kubota et al., 1999) they proposed a two-stage registration method based on hierarchical matching, where the amount of blur is considered as another parameter of the search space. In (Zhang and Blum, 2001) they proposed an iterative multiscale registration based on optical flow estimation in each scale, claiming that optical flow estimation is robust to image blurring. All global methods require considerable (or even complete) spatial overlap of the channels to yield reliable results, which is their major drawback.

Landmark-based blur-invariant registration methods have appeared very recently, just after the first paper on the moment-based blur-invariant features (Flusser et al., 1996). Originally, these features could only be used for registration of mutually shifted images (Flusser and Suk, 1998), (Bentoutou et al., 2002). The proposal of their rotational-invariant version (Flusser and Zitová, 1999) in combination with a robust detector of salient points (Zitová et al., 1999) led to the registration methods that are able to handle blurred, shifted and rotated images (Flusser et al., 1999), (Flusser et al., 2003).

Although the above-cited registration methods are very sophisticated and can be applied almost to all types of images, the results tend to be rarely perfect. The registration error usually varies from subpixel values to a few pixels, so only MBD and SR methods sufficiently robust to between-channel misregistration can be applied to channel fusion. We will assume in the sequel that the LR images are roughly registered and that W_k's reduce to small translations.

During the last twenty years, blind deconvolution has attracted considerable attention as a separate image processing task. Initial blind deconvolution attempts were based on single-channel formulations, such as in (Lagendijk et al., 1990; Reeves and Mersereau, 1992; Chan and Wong, 1998; Haindl, 2000). A good overview is in (Kundur and Hatzinakos, 1996a; Kundur and Hatzinakos, 1996b). The problem is extremely ill-posed in the single-channel framework and cannot be resolved in the fully blind form. These methods do not exploit the potential of multiframe imaging, because in the single-channel case the missing information about the original image in one channel cannot by supplemented by information obtained from the other channels. Research on intrinsically multichannel methods has begun fairly recently; refer to (Harikumar and Bresler, 1999; Giannakis and Heath, 2000; Pai and Bovik, 2001; Panci et al., 2003; Šroubek and Flusser, 2003) for a survey and other references. Such MBD methods break the limitations of previous techniques and can recover the blurring functions from the degraded images alone. We further developed the MBD theory in (Šroubek and Flusser, 2005) by proposing a blind deconvolution method for images, which might be mutually shifted by unknown vectors. A similar idea is used here as a part of the fusion algorithm to remove volatile blurs and will be explained more in Section 3.

Superresolution has been mentioned in the literature with an increasing frequency in the last decade. The first SR methods did not involve any deblurring; they just tried to register the LR images with subpixel accuracy and then to resample them on a high-resolution grid. A good survey of SR techniques can be found in (Park et al., 2003; Farsui et al., 2004). Maximum likelihood (ML), maximum a posteriori (MAP), the set theoretic approach using POCS (projection on convex sets), and fast Fourier techniques can all provide a solution to the SR problem. Earlier approaches assumed that subpixel shifts are estimated by other means. More advanced techniques, such as in (Hardie et al., 1997; Segall et al., 2004; Woods et al., 2006), include shift estimation in the SR process. Other approaches focus on fast implementation (Farsiu et al., 2004), space-time SR (Shechtman et al., 2005) or SR of compressed video (Segall et al., 2004). Some of the recent SR methods consider image blurring and involve blur removal. Most of them assume only *a priori* known blurs. However, few exceptions exist. Authors in (Nguyen et al., 2001; Woods et al., 2003) proposed BSR that can handle parametric PSFs with one parameter. This restriction is unfortunately very limiting for most real applications. Probably the first attempts for BSR with an arbitrary PSF appeared in (Wirawan et al., 1999; Yagle, 2003), where polyphase decomposition of the images was employed.

Current multiframe blind deconvolution techniques require no or very little prior information about the blurs, they are sufficiently robust to noise and provide satisfying results in most real applications. However, they can hardly cope with the downsampling operator, which violates the standard convolution model. On the contrary, state-of-the-art SR techniques achieve remarkable results in resolution enhancement in the case of no blur. They accurately estimate the subpixel shift between images but lack any apparatus for calculating the blurs.

We propose a unifying method that simultaneously estimates the volatile blurs and HR image without any prior knowledge of the blurs and the original image. We accomplish this by formulating the problem as a minimization of a regularized energy function, where the regularization is carried out in both the image and blur domains. Image regularization is based on variational integrals, and a consequent anisotropic diffusion with good edge-preserving capabilities. A typical example of such regularization is total variation. However, the main contribution of this work lies in the development of the blur regularization term. We show that the blurs can be recovered from the LR images up to small ambiguity. One can consider this as a generalization of the results proposed for blur estimation in the case of MBD problems. This fundamental observation enables us to build a simple regularization term for the blurs even in the case of the SR problem. To tackle the minimization task we use an alternating minimization approach, consisting of two simple linear equations.

The rest of the chapter is organized as follows. Section 2 outlines the degradation model. In Section 3 we present a procedure for volatile blur estimation. This effortlessly blends in a regularization term of the BSR algorithm as described in

Section 4. Finally, Section 5 illustrates applicability of the proposed method to real situations.

2. Mathematical Model

To simplify the notation, we will assume only images and PSFs with square supports. An extension to rectangular images is straightforward. Let $f[x,y]$ be an arbitrary discrete image of size $F \times F$, then $\mathbf{f}$ denotes an image column vector of size $F^2 \times 1$ and $\mathbf{C}_A\{f\}$ denotes a matrix that performs convolution of f with an image of size $A \times A$. The convolution matrix can have a different output size. Adopting the Matlab naming convention, we distinguish two cases: "full" convolution $\mathbf{C}_A\{f\}$ of size $(F + A - 1)^2 \times A^2$ and "valid" convolution $\mathbf{C}_A^v\{f\}$ of size $(F - A + 1)^2 \times A^2$. In both cases the convolution matrix is a Toeplitz-block-Toeplitz (TBT) matrix. In the sequel we will not specify dimensions of convolution matrices if it is obvious from the size of the right argument.

Let us assume we have K different LR frames $\{g_k\}$ (each of size $G \times G$) that represent degraded (blurred and noisy) versions of the original scene. Our goal is to estimate the HR representation of the original scene, which we denoted as the HR image f of size $F \times F$. The LR frames are linked with the HR image through a series of degradations similar to those between $o[x,y]$ and g_k in (2). First f is geometrically warped ($\mathbf{W}_k$), then it is convolved with a volatile PSF ($\mathbf{V}_k$) and finally it is decimated ($\mathbf{D}$). The formation of the LR images in vector-matrix notation is then described as

$$\mathbf{g}_k = \mathbf{D}\mathbf{V}_k\mathbf{W}_k\mathbf{f} + \mathbf{n}_k\,, \tag{3}$$

where $\mathbf{n}_k$ is additive noise present in every channel. The decimation matrix $\mathbf{D} = \mathbf{SU}$ simulates the behavior of digital sensors by first performing convolution with the $U \times U$ sensor PSF ($\mathbf{U}$) and then downsampling ($\mathbf{S}$). The Gaussian function is widely accepted as an appropriate sensor PSF and it is also used here. Its justification is experimentally verified in (Capel, 2004). A physical interpretation of the sensor blur is that the sensor is of finite size and it integrates impinging light over its surface. The sensitivity of the sensor is highest in the middle and decreases towards its borders with Gaussian-like decay. Further we assume that the subsampling factor (or SR factor, depending on the point of view), denoted by ε, is the same in both x and y directions. It is important to underline that ε is a user-defined parameter. In principle, $\mathbf{W}_k$ can be a very complex geometric transform that must be estimated by image registration or motion detection techniques. We have to keep in mind that sub-pixel accuracy in $\mathbf{g}_k$'s is necessary for SR to work. Standard image registration techniques can hardly achieve this and they leave a small misalignment behind. Therefore, we will assume that complex geometric transforms are removed in the preprocessing step and $\mathbf{W}_k$ reduces to a small translation. Hence $\mathbf{V}_k\mathbf{W}_k = \mathbf{H}_k$, where $\mathbf{H}_k$ performs convolution with the shifted version of the volatile PSF v_k, and the acquisition model becomes

$$\mathbf{g}_k = \mathbf{D}\mathbf{H}_k\mathbf{f} + \mathbf{n}_k = \mathbf{S}\mathbf{U}\mathbf{H}_k\mathbf{f} + \mathbf{n}_k\,. \tag{4}$$

The BSR problem then adopts the following form: We know the LR images $\{g_k\}$ and we want to estimate the HR image f for the given $\mathbf{S}$ and the sensor blur $\mathbf{U}$. To avoid boundary effects, we assume that each observation g_k captures only a part of f. Hence $\mathbf{H}_k$ and $\mathbf{U}$ are “valid” convolution matrices $\mathbf{C}^v_F\{h_k\}$ and $\mathbf{C}^v_{F-H+1}\{u\}$, respectively. In general, the PSFs h_k are of different size. However, we postulate that they all fit into a $H \times H$ support.

In the case of $\varepsilon = 1$, the downsampling $\mathbf{S}$ is not present and we face a slightly modified MBD problem that has been solved elsewhere (Harikumar and Bresler, 1999; Šroubek and Flusser, 2005). Here we are interested in the case of $\varepsilon > 1$, when the downsampling occurs. Can we estimate the blurs as in the case $\varepsilon = 1$? The presence of $\mathbf{S}$ prevents us from using the cited results directly. However, we will show that conclusions obtained for MBD apply here in a slightly modified form as well.

3. Reconstruction of Volatile Blurs

Estimation of blurs in the MBD case (no downsampling) attracted considerable attention in the past. A wide variety of methods were proposed, such as in (Harikumar and Bresler, 1999; Giannakis and Heath, 2000), that provide a satisfactory solution. For these methods to work correctly, certain channel disparity is necessary. The disparity is defined as weak co-primeness of the channel blurs, which states that the blurs have no common factor except a scalar constant. In other words, if the channel blurs can be expressed as a convolution of two subkernels then there is no subkernel that is common to all blurs. An exact definition of weakly co-prime blurs can be found in (Giannakis and Heath, 2000). Many practical cases satisfy the channel co-primeness, since the necessary channel disparity is mostly guaranteed by the nature of the acquisition scheme and random processes therein. We refer the reader to (Harikumar and Bresler, 1999) for a relevant discussion. This channel disparity is also necessary for the BSR case.

Let us first recall how to estimate blurs in the MBD case and then we will show how to generalize the results for integer downsampling factors. For the time being we will omit noise n, until Section 4, where we will address it appropriately.

3.1. THE MBD CASE

The downsampling matrix $\mathbf{S}$ is not present in (4) and only convolution binds the input with the outputs. The acquisition model is of the SIMO type with one input channel f and K output channels g_k. Under the assumption of channel co-primeness, we can see that any two correct blurs h_i and h_j satisfy

$$\|g_i * h_j - g_j * h_i\|^2 = 0\,. \tag{5}$$

Considering all possible pairs of blurs, we can arrange the above relation into one system

$$\mathcal{N}'\mathbf{h} = \mathbf{0}\,, \tag{6}$$

where $\mathbf{h} = [\mathbf{h}_1^T, \ldots, \mathbf{h}_K^T]^T$ and $\mathcal{N}'$ consists of matrices that perform convolution with g_k. In most real situations the correct blur size (we have assumed square size $H \times H$) is not known in advance and therefore we can generate the above equation for different blur dimensions $\hat{H}_1 \times \hat{H}_2$. The nullity (null-space dimension) of $\mathcal{N}'$ is exactly 1 for the correctly estimated blur size. By applying SVD (singular value decomposition), we recover precisely the blurs except for a scalar factor. One can eliminate this magnitude ambiguity by stipulating that $\sum_{x,y} h_k[x,y] = 1$, which is a common brightness preserving assumption. For the underestimated blur size, the above equation has no solution. If the blur size is overestimated, then nullity$(\mathcal{N}') = (\hat{H}_1 - H + 1)(\hat{H}_2 - H + 1)$.

3.2. THE BSR CASE

Before we proceed, it is necessary to define precisely the sampling matrix $\mathbf{S}$. Let $\mathbf{S}_1^{\varepsilon}$ denote a 1-D sampling matrix, where ε is the integer subsampling factor. Each row of the sampling matrix is a unit vector whose nonzero element is at such a position that, if the matrix multiplies an arbitrary vector b, the result of the product is every ε-th element of b starting from b_1. If the vector length is M then the size of the sampling matrix is $(M/\varepsilon) \times M$. If M is not divisible by ε, we can pad the vector with an appropriate number of zeros to make it divisible. A 2-D sampling matrix is defined by

$$\mathbf{S}^{\varepsilon} := \mathbf{S}_1^{\varepsilon} \otimes \mathbf{S}_1^{\varepsilon}\,, \tag{7}$$

where $\otimes$ denotes the matrix direct product (Kronecker product operator). Note that the transposed matrix $(\mathbf{S}^{\varepsilon})^T$ behaves as an upsampling operator that interlaces the original samples with $(\varepsilon - 1)$ zeros.

A naive approach, e.g. proposed in (Šroubek and Flusser, 2006; Chen et al., 2005), is to modify (6) in the MBD case by applying downsampling and formulating the problem as

$$\min_{\mathbf{h}} \|\mathcal{N}'[\mathbf{I}_K \otimes \mathbf{S}^{\varepsilon}\mathbf{U}]\mathbf{h}\|^2\,, \tag{8}$$

where $\mathbf{I}_K$ is the $K \times K$ identity matrix. One can easily verify that the condition in (5) is not satisfied for the BSR case as the presence of downsampling operators violates the commutative property of convolution. Even more disturbing is the fact that minimizers of (8) do not have to correspond to the correct blurs. We are going to show that if one uses a slightly different approach, reconstruction of the volatile PSFs h_k is possible even in the BSR case. However, we will see that some ambiguity in the solution of h_k is inevitable.

First, we need to rearrange the acquisition model (4) and construct from the LR images g_k a convolution matrix $\mathcal{G}$ with a predetermined nullity. Then we take the null space of $\mathcal{G}$ and construct a matrix $\mathcal{N}$, which will contain the correct PSFs h_k in its null space.

Let $E \times E$ be the size of "nullifying" filters. The meaning of this name will be clear later. Define $\mathcal{G} := [\mathbf{G}_1, \ldots, \mathbf{G}_K]$, where $\mathbf{G}_k := \mathbf{C}_E^v\{g_k\}$ are "valid" convolution matrices. Assuming no noise, we can express $\mathcal{G}$ in terms of f, u and h_k as

$$\mathcal{G} = \mathbf{S}^\varepsilon \mathbf{F} \mathbf{U} \mathcal{H}\,, \tag{9}$$

where

$$\mathcal{H} := [\mathbf{C}_{\varepsilon E}\{h_1\}(\mathbf{S}^\varepsilon)^T, \ldots, \mathbf{C}_{\varepsilon E}\{h_K\}(\mathbf{S}^\varepsilon)^T]\,, \tag{10}$$

$\mathbf{U} := \mathbf{C}_{\varepsilon E+H-1}\{u\}$ and $\mathbf{F} := \mathbf{C}^v_{\varepsilon E+H+U-2}\{f\}$.

The convolution matrix $\mathcal{U}$ has more rows than columns and therefore it is of full column rank (see proof in (Harikumar and Bresler, 1999) for general convolution matrices). We assume that $\mathbf{S}^\varepsilon\mathbf{F}$ has full column rank as well. This is almost certainly true for real images if $\mathbf{F}$ has at least ε^2-times more rows than columns. Thus $\text{Null}(\mathcal{G}) \equiv \text{Null}(\mathcal{H})$ and the difference between the number of columns and rows of $\mathcal{H}$ bounds from below the null space dimension, i.e.,

$$\text{nullity}(\mathcal{G}) \geq KE^2 - (\varepsilon E + H - 1)^2\,. \tag{11}$$

Setting $N := KE^2 - (\varepsilon E + H - 1)^2$ and $\mathbf{N} := \text{Null}(\mathcal{G})$, we visualize the null space as

$$\mathbf{N} = \begin{bmatrix} \mathbf{n}_{1,1} & \ldots & \mathbf{n}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{n}_{K,1} & \ldots & \mathbf{n}_{K,N} \end{bmatrix}, \tag{12}$$

where $\mathbf{n}_{kn}$ is the vector representation of the nullifying filter η_{kn} of size $E \times E$, $k = 1, \ldots, K$ and $n = 1, \ldots, N$. Let $\tilde{\eta}_{kn}$ denote upsampled η_{kn} by factor ε, i.e., $\tilde{\eta}_{kn} := (\mathbf{S}^\varepsilon)^T \eta_{kn}$. Then, we define

$$\mathcal{N} := \begin{bmatrix} \mathbf{C}_H\{\tilde{\eta}_{1,1}\} & \ldots & \mathbf{C}_H\{\tilde{\eta}_{K,1}\} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_H\{\tilde{\eta}_{1,N}\} & \ldots & \mathbf{C}_H\{\tilde{\eta}_{K,N}\} \end{bmatrix} \tag{13}$$

and conclude that

$$\mathcal{N}\mathbf{h} = \mathbf{0}\,, \tag{14}$$

where $\mathbf{h}^T = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$. We have arrived at an equation that is of the same form as (6) in the MBD case. Here we have the solution to the blur estimation problem for the BSR case. However, since $\mathbf{S}^\varepsilon$ is involved, ambiguity of the solution is higher. Without proofs we provide the following statements. For the correct blur size, nullity$(\mathcal{N}) = \varepsilon^4$. For the underestimated blur size, (14) has no solution. For the overestimated blur size $\hat{H}_1 \times \hat{H}_2$, nullity$(\mathcal{N}) = \varepsilon^2(\hat{H}_1 - H + \varepsilon)(\hat{H}_2 - H + \varepsilon)$.

The conclusion may seem to be pessimistic. For example, for $\varepsilon = 2$ the nullity is at least 16, and for $\varepsilon = 3$ the nullity is already 81. Nevertheless, Section 4 will show that $\mathcal{N}$ plays an important role in the regularized restoration algorithm and its ambiguity is not a serious drawback.

It is interesting to note that a similar derivation is possible for rational SR factors $\varepsilon = p/q$. We downsample the LR images with the factor q, thereby creating q^2K images, and apply thereon the above procedure for the SR factor p.

Another consequence of the above derivation is the minimum necessary number of LR images for the blur reconstruction to work. The condition of the $\mathcal{G}$ nullity in (11) implies that the minimum number is $K > \varepsilon^2$. For example, for $\varepsilon = 3/2$, 3 LR images are sufficient; for $\varepsilon = 2$, we need at least 5 LR images to perform blur reconstruction.

4. Blind Superresolution

In order to solve the BSR problem, i.e, determine the HR image f and volatile PSFs h_k, we adopt a classical approach of minimizing a regularized energy function. This way the method will be less vulnerable to noise and better posed. The energy consists of three terms and takes the form

$$E(\mathbf{f}, \mathbf{h}) = \sum_{k=1}^{K} ||\mathbf{D}\mathbf{H}_k\mathbf{f} - \mathbf{g}_k||^2 + \alpha Q(\mathbf{f}) + \beta R(\mathbf{h})\,. \tag{15}$$

The first term measures the fidelity to the data and emanates from our acquisition model (4). The remaining two are regularization terms with positive weighting constants α and β that attract the minimum of E to an admissible set of solutions. The form of E very much resembles the energy proposed in (Šroubek and Flusser, 2005) for MBD. Indeed, this should not come as a surprise since MBD and SR are related problems in our formulation.

Regularization $Q(\mathbf{f})$ is a smoothing term of the form

$$Q(\mathbf{f}) = \mathbf{f}^T\mathbf{L}\mathbf{f}\,, \tag{16}$$

where $\mathbf{L}$ is a high-pass filter. A common strategy is to use convolution with the Laplacian for $\mathbf{L}$, which in the continuous case corresponds to $Q(f) = \int |\nabla f|^2$. Recently, variational integrals $Q(f) = \int \phi(|\nabla f|)$ were proposed, where ϕ is a strictly convex, nondecreasing function that grows at most linearly. Examples of $\phi(s)$ are s (total variation), $\sqrt{1+s^2} - 1$ (hypersurface minimal function), $\log(\cosh(s))$, or nonconvex functions, such as $\log(1+s^2)$, $s^2/(1+s^2)$ and $\arctan(s^2)$ (Mumford-Shah functional). The advantage of the variational approach is that, while in smooth areas it has the same isotropic behavior as the Laplacian, it also preserves edges in images. The disadvantage is that it is highly nonlinear. To overcome this difficulty one must use, e.g., the half-quadratic algorithm (Aubert and Kornprobst, 2002). For the purpose of our discussion it suffices to state that after discretization we arrive again at (16), where this time $\mathbf{L}$ is a positive semidefinite block tridiagonal matrix constructed of values depending on the gradient of f. The rationale behind the choice of $Q(f)$ is to constrain the local spatial behavior of images; it resembles a Markov Random Field. Some global constraints may be more desirable but are difficult (often impossible) to define, since we develop a general method that should work with any class of images.

The PSF regularization term $R(\mathbf{h})$ directly follows from the conclusions of the previous section. Since the matrix $\mathcal{N}$ in (13) contains the correct PSFs h_k in its null space, we define the regularization term as a least-squares fit

$$R(\mathbf{h}) = \|\mathcal{N}\mathbf{h}\|^2 = \mathbf{h}^T\mathcal{N}^T\mathcal{N}\mathbf{h}\,. \tag{17}$$

The product $\mathcal{N}^T\mathcal{N}$ is a positive semidefinite matrix. More precisely, R is a consistency term that binds the different volatile PSFs to prevent them from moving freely and, unlike the fidelity term (the first term in (15)), it is based solely on the observed LR images. A good practice is to include with a small weight a smoothing term $\mathbf{h}^T\mathbf{L}\mathbf{h}$ in $R(\mathbf{h})$. This is especially useful in the case of less noisy data to overcome the higher nullity of $\mathcal{N}$.

The complete energy then takes the form

$$E(\mathbf{f},\mathbf{h}) = \sum_{k=1}^{K} \|\mathbf{D}\mathbf{H}_k\mathbf{f} - \mathbf{g}_k\|^2 + \alpha\mathbf{f}^T\mathbf{L}\mathbf{f} + \beta_1\|\mathcal{N}\mathbf{h}\|^2 + \beta_2\mathbf{h}^T\mathbf{L}\mathbf{h}\,. \tag{18}$$

To find a minimizer of the energy function, we perform alternating minimizations (AM) of E over $\mathbf{f}$ and $\mathbf{h}$. The advantage of this scheme lies in its simplicity. Each term of (18) is quadratic and therefore convex (but not necessarily strictly convex) and the derivatives w.r.t. $\mathbf{f}$ and $\mathbf{h}$ are easy to calculate. This AM approach is a variation on the steepest-descent algorithm. The search space is a concatenation of the blur subspace and the image subspace. The algorithm first descends in the image subspace and after reaching the minimum, i.e., $\nabla_{\mathbf{f}}E = 0$, it advances in the blur subspace in the direction $\nabla_{\mathbf{h}}E$ orthogonal to the previous one, and this scheme repeats. In conclusion, starting with some initial $\mathbf{h}^0$ the two iterative steps are:

$$\begin{aligned} \text{step 1.} \qquad \mathbf{f}^m = {} & \arg\min_{\mathbf{f}} E(\mathbf{f},\mathbf{h}^m) \\ \Leftrightarrow \quad & \Big(\sum_{k=1}^{K} \mathbf{H}_k^T\mathbf{D}^T\mathbf{D}\mathbf{H}_k + \alpha\mathbf{L}\Big)\mathbf{f} = \sum_{k=1}^{K} \mathbf{H}_k^T\mathbf{D}^T\mathbf{g}_k\,, \end{aligned} \tag{19}$$

$$\text{step 2.}\quad \mathbf{h}^{m+1} = \arg\min_{\mathbf{h}} E(\mathbf{f}^m, \mathbf{h})$$

$$\Leftrightarrow \quad ([\mathbf{I}_K \otimes \mathbf{F}^T \mathbf{D}^T \mathbf{D} \mathbf{F}] + \beta_1 \mathcal{N}^T \mathcal{N} + \beta_2 \mathbf{L})\mathbf{h} = [\mathbf{I}_K \otimes \mathbf{F}^T \mathbf{D}^T]\mathbf{g} \,, \quad (20)$$

where $\mathbf{F} := \mathbf{C}^v_H\{f\}$, $\mathbf{g} := [\mathbf{g}_1^T, \ldots, \mathbf{g}_K^T]^T$ and m is the iteration step. Note that both steps consist of simple linear equations.

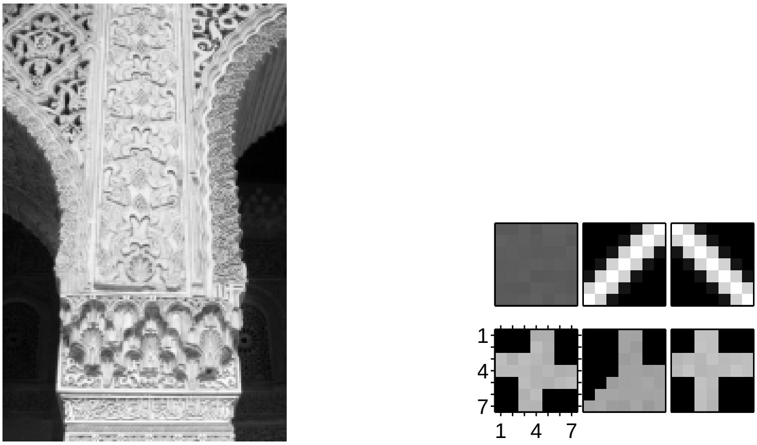
Energy E as a function of both variables $\mathbf{f}$ and $\mathbf{h}$ is not convex due to the coupling of the variables via convolution in the first term of (18). Therefore, it is not guaranteed that the BSR algorithm reaches the global minimum. In our experience, convergence properties improve significantly if we add feasible regions for the HR image and PSFs specified as lower and upper bounds constraints. To solve step 1, we use the method of conjugate gradients (function *cgs* in Matlab) and then adjust the solution $\mathbf{f}^m$ to contain values in the admissible range, typically, the range of values of $\mathbf{g}$. It is common to assume that PSF is positive ($h_k \geq 0$) and that it preserves image brightness. We can therefore write the lower and upper bounds constraints for PSFs as $\mathbf{h}_k \in \langle 0, 1\rangle^{H^2}$. In order to enforce the bounds in step 2, we solve (20) as a constrained minimization problem (function *fmincon* in Matlab) rather than using the projection as in step 1. Constrained minimization problems are more computationally demanding but we can afford it in this case since the size of $\mathbf{h}$ is much smaller than the size of $\mathbf{f}$.

The weighting constants α and β_i depend on the level of noise. If noise increases, α and β_2 should increase, and β_1 should decrease. One can use parameter estimation techniques, such as cross-validation (Nguyen et al., 2001) or expectation maximization (Molina et al., 2003), to determine the correct weights. However, in our experiments we set the values manually according to a visual assessment. If the iterative algorithm begins to amplify noise, we have underestimated the noise level. On the contrary, if the algorithm begins to segment the image, we have overestimated the noise level.

5. Experiments

This section consists of two parts. In the first one, a set of experiments on synthetic data evaluate performance of the BSR algorithm with respect to the SR factor and compare the reconstruction quality with other methods. The second part demonstrates the applicability of the proposed method to real data. Results are not evaluated with any measure of reconstruction quality, such as mean-square errors or peak signal to noise ratios. Instead we print the results and leave the comparison to a human eye as we believe that in this case the visual assessment is the only reasonable method.

In all the experiments the sensor blur is fixed and set to a Gaussian function of standard deviation $\sigma = 0.34$ (relative to the scale of LR images). One should underline that the proposed BSR method is fairly robust to the choice of the Gaus-



a. b.

Figure 2. Simulated data: (a) original 150 × 230 image; (b) six 7 × 7 volatile PSFs used to blur the original image.

sian variance, since it can compensate for insufficient variance by automatically including the missing factor of Gaussian functions in the volatile blurs.

Another potential pitfall that we have to take into consideration is a feasible range of SR factors. Clearly, as the SR factor ε increases we need more LR images and the stability of BSR decreases. In addition, rational SR factors p/q, where p and q are incommensurable and large regardless of the effective value of ε, also make the BSR algorithm unstable. It is the numerator p that determines the internal SR factor used in the algorithm. Hence we limit ourselves to ε between 1 and 2.5, such as 3/2, 5/3, 2, etc., which is sufficient in most practical applications.

5.1. SIMULATED DATA

First, let us demonstrate the BSR performance with a simple experiment. A 150 × 230 image in Fig. 2.a blurred with the six masks in Fig. 2.b and downsampled with factor 2 generated six LR images. In this case, registration is not necessary since the synthetic data are precisely aligned. Using the LR images as an input, we estimated the original HR image with the proposed BSR algorithm for $\varepsilon = 1.25$ and 1.75. In Fig. 3 one can compare the results printed in their original size. The HR image for $\varepsilon = 1.25$ (Fig. 3.b) has improved significantly on the LR images due to deconvolution, however some details on the column are still distorted. For the SR factor 1.75, the reconstructed image in Fig. 3.c is almost perfect.

Next we compare performance of the BSR algorithm with two methods: interpolation technique and state-of-the-art SR method. The former technique consists of the MBD method proposed in (Šroubek and Flusser, 2005) followed by

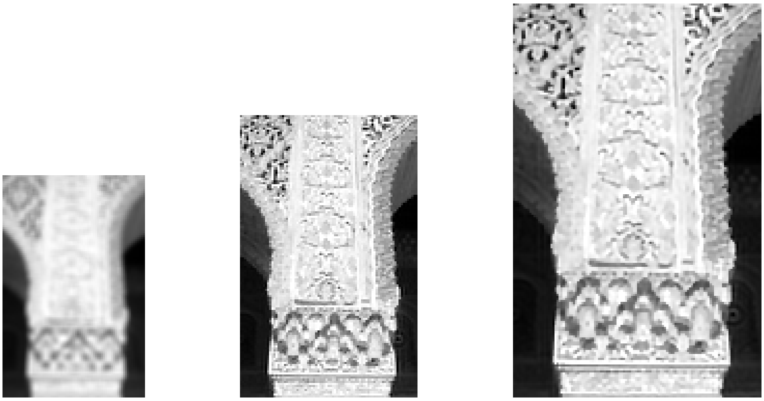

a. LR b. $\varepsilon = 1.25$ c. $\varepsilon = 1.75$

Figure 3. BSR of simulated data: (a) one of six LR images with the downsampling factor 2; (b) BSR for $\varepsilon = 1.25$; (c) BSR for $\varepsilon = 1.75$.

standard bilinear interpolation (BI) resampling. The MBD method first removes volatile blurs and then BI of the deconvolved image achieves the desired spatial resolution. The latter method, which we will call herein a “standard SR algorithm”, is a MAP formulation of the SR problem proposed, e.g., in (Hardie et al., 1997; Segall et al., 2004). This method uses a MAP framework for the joint estimation of image registration parameters (in our case only translation) and the HR image, assuming only the sensor blur (**U**) and no volatile blurs. For an image prior, we use edge preserving Huber Markov Random Fields (Capel, 2004).

In the case of BSR, Section 3 has shown that two distinct approaches exist for blur estimation. Either we use the naive approach in (8) that directly utilizes the MBD formulation, or we apply the intrinsically SR approach summarized in (14). Altogether we have thus four distinct methods for comparison: standard SR approach, MBD with interpolation, BSR with naive blur regularization and BSR with intrinsic blur regularization. Using the original image and PSFs in Fig. 2, six LR images (see one LR image in Fig. 3.a) were generated as in the first experiment, only this time we added white Gaussian noise with SNR = 50dB[1].

Estimated HR images and volatile blurs for all four methods are in Fig. 4. The standard SR approach in Fig. 4.a gives unsatisfactory results, since heavy blurring is present in the LR images and the method assumes only the sensor blur and no volatile blurs. (For this reason, we do not show volatile blurs in this case). The MBD method in Fig. 4.b ignores the decimation operator and thus the estimated volatile blurs are similar to LR projections of the original blurs. Despite the fact that blind deconvolution in the first stage performed well, many

[1] The signal-to-noise ratio is defined as $\mathrm{SNR} = 10\log(\sigma_f^2/\sigma_n^2)$, where σ_f and σ_n are the image and noise standard deviations, respectively.

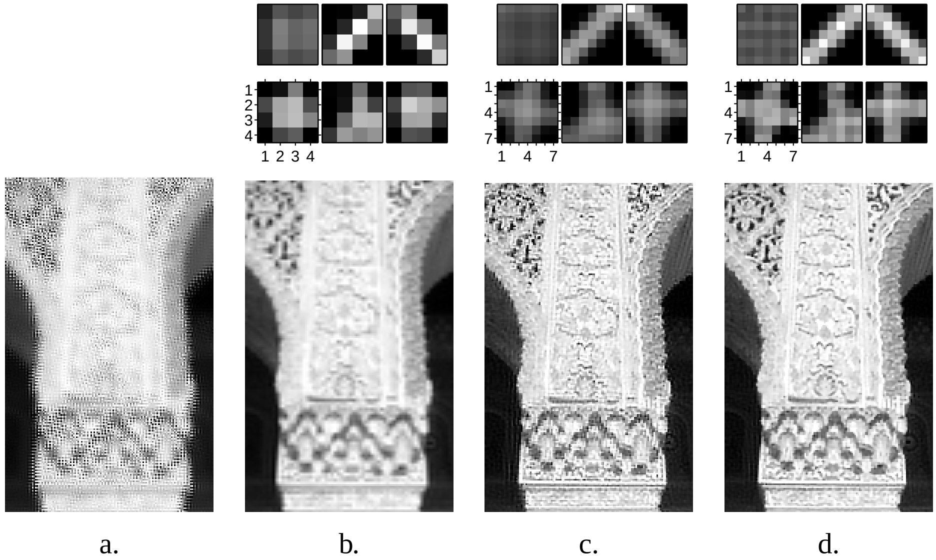


Figure 4. Comparison of four different SR approaches ($\varepsilon = 2$): (a) standard SR method, (b) MBD followed by bilinear interpolation, (c) naive BSR approach and (b) proposed intrinsic BSR approach. Volatile blurs estimated by each method, except in the case of standard SR, are in the top row. Due to blurring, the standard SR method in (a) failed to reconstruct the HR image. MBD in (b) provided a good estimate of the blurs in the LR scale and performed correct deconvolution but the HR image lacks many details as simple interpolation increased resolution. Both BSR approaches in (c) and (d) gave close to perfect results. However in the case of the naive approach, inaccurate blur regularization resulted in several artifacts in the HR image.

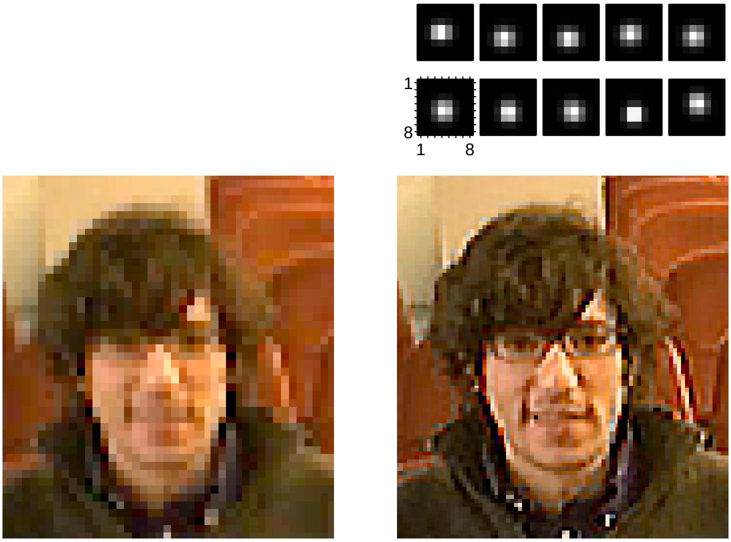
details are still missing since interpolation in the second stage cannot properly recover high-frequency information. Both the naive and the intrinsic BSR methods outperformed the previous approaches and the intrinsic one provides a close-to-perfect HR image. Due to the inaccurate regularization term in the naive approach, estimated blurs contain tiny erroneous components that resulted in artifacts in the HR image (Fig. 4.c). However, a more strict and accurate regularization term in the case of the intrinsic BSR approach improved results, which one can see in Fig. 4.d.

5.2. REAL DATA

The next two experiments demonstrate the true power of our fusion algorithm. We used real photos acquired with two different acquisition devices: webcamera and standard digital camera. The webcam was Logitech QuickCam for Notebooks Pro with the maximum video resolution 640 × 480 and the minimum shutter speed 1/10s. The digital camera was 5 Mpixel Olympus C5050Z equipped with 3×

optical zoom. In both experiments we used cross-correlation to roughly register the LR images.

In the first one we hold the webcam in hands and captured a short video sequence of a human face. Then we extracted 10 consecutive frames and considered a small section of size 40 × 50. One frame with zero-order interpolation is in Fig. 5.a. The other frames look similar. The long shutter speed (1/10s) together with the inevitable motion of hands introduced blurring into the images. In this experiment, the SR factor was set to 2. The proposed BSR algorithm removed blurring and performed SR correctly as one can see in Fig. 5.b. Note that many facial features (eyes, glasses, mouth) indistinguishable in the original LR image became visible in the HR image.



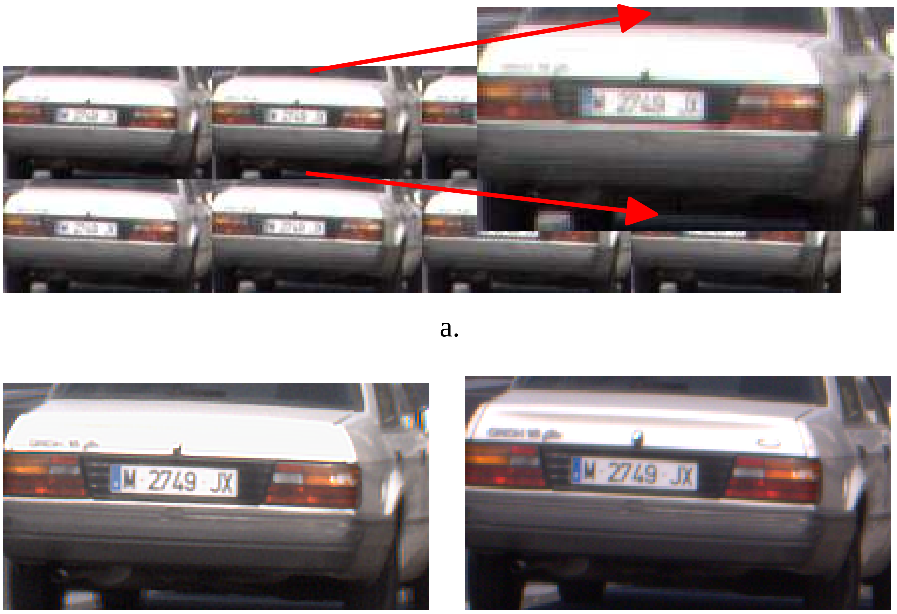
a.

b.

Figure 5. Reconstruction of images acquired with a webcam ($\varepsilon = 2$): (a) one of ten LR frames extracted from a short video sequence captured with the webcam, zero-order interpolation; (b) HR image and blurs estimated by the BSR algorithm. Note that many facial features, such as glasses, are not apparent in the LR image, but are well reconstructed in the HR image.

The second experiment demonstrates a task of license plate recognition. With the digital camera we took eight photos, registered them with cross-correlation and cropped each to a 100×50 rectangle. All eight cuttings printed in their original size (no interpolation), including one image enlarged with zero-order interpolation, are in Fig. 6.a. Similar to the previous experiment, the camera was held in hands, and due to the longer shutter speed, the LR images exhibit subtle blurring. We set the SR factor to 5/3. In order to better assess the obtained results we took one additional image with optical zoom 1.7× (close to the desired SR factor 5/3).

This image served as the ground truth; see Fig. 6.c. The proposed BSR method returned a well reconstructed HR image (Fig. 6.b), which is comparable to the ground truth acquired with the optical zoom.



a.

b.

c.

Figure 6. Reconstruction of images acquired with a digital camera ($\varepsilon = 5/3$): (a) eight LR images, one enlarged with zero-order interpolation; (b) HR image estimated by the BSR algorithm; (c) image acquired with optical zoom 1.7×. The BSR algorithm achieved reconstruction comparable to the image with optical zoom.

6. Conclusions

In this chapter we proposed a method for improving visual quality and spatial resolution of digital images acquired by low-resolution sensors. The method is based on fusing several images (channels) of the same scene. It consists of three major steps – image registration, blind deconvolution and superresolution enhancement. We reviewed all three steps and we paid special attention to superresolution fusion. We proposed a unifying system that simultaneously estimates image blurs and recovers the original undistorted image, all in high resolution, without any prior knowledge of the blurs and original image. We accomplished this by formulating the problem as constrained least squares energy minimization with appropriate regularization terms, which guarantees a close-to-perfect solution.

Showing the good performance of the method on real data, we demonstrated its capability to improve the image quality significantly and, consequently, to make the task of object detection and identification much easier for human observers as well as for automatic systems. We envisage the application of the proposed method in security and surveillance systems.

Acknowledgements

This work has been supported by Czech Ministry of Education under project No. 1M0572 (Research Center DAR), and by the Grant Agency of the Czech Republic under project No. 102/04/0155.

References

Aubert, G. and Kornprobst, P. (2002) *Mathematical Problems in Image Processing,* New York, Springer Verlag.

Bentoutou, Y., Taleb, N., Mezouar, M., Taleb, M., and Jetto, L. (2002) An invariant approach for image registration in digital subtraction angiography, *Pattern Recognition* **35**, 2853–2865.

Capel, D. (2004) *Image Mosaicing and Super-Resolution,* New York, Springer.

Chan, T. and Wong, C. (1998) Total variation blind deconvolution, *IEEE Trans. Image Processing* **7**, 370–375.

Chen, Y., Luo, Y., and Hu, D. (2005) A general approach to blind image super-resolution using a PDE framework, In *Proc. SPIE,* Vol. 5960, pp. 1819–1830.

Farsiu, S., Robinson, M., Elad, M., and Milanfar, P. (2004) Fast and robust multiframe super resolution, *IEEE Trans. Image Processing* **13**, 1327–1344.

Farsui, S., Robinson, D., Elad, M., and Milanfar, P. (2004) Advances and challenges in super-resolution, *Int. J. Imag. Syst. Technol.* **14**, 47–57.

Flusser, J., Boldyš, J., and Zitová, B. (2003) Moment forms invariant to rotation and blur in arbitrary number of dimensions, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 234–246.

Flusser, J. and Suk, T. (1998) Degraded image analysis: an invariant approach, *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**, 590–603.

Flusser, J., Suk, T., and Saic, S. (1996) Recognition of blurred images by the method of moments, *IEEE Trans. Image Processing* **5**, 533–538.

Flusser, J. and Zitová, B. (1999) Combined invariants to linear filtering and rotation, *Intl. J. Pattern Recognition Art. Intell.* **13**, 1123–1136.

Flusser, J., Zitová, B., and Suk, T. (1999) Invariant-based registration of rotated and blurred images, In I. S. Tammy (ed.), *Proceedings IEEE 1999 International Geoscience and Remote Sensing Symposium*, Los Alamitos, pp. 1262–1264, IEEE Computer Society.

Giannakis, G. and Heath, R. (2000) Blind identification of multichannel FIR blurs and perfect image restoration, *IEEE Trans. Image Processing* **9**, 1877–1896.

Haindl, M. (2000) Recursive model-based image restoration, In *Proceedings of the 15th International Conference on Pattern Recognition*, Vol. III, pp. 346–349, IEEE Press.

Hardie, R., Barnard, K., and Armstrong, E. (1997) Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, *IEEE Trans. Image Processing* **6**, 1621–1633.

Harikumar, G. and Bresler, Y. (1999) Perfect blind restoration of images blurred by multiple filters: theory and efficient algorithms, *IEEE Trans. Image Processing* **8**, 202–219.

Kubota, A., Kodama, K., and Aizawa, K. (1999) Registration and blur estimation methods for multiple differently focused images, In *Proceedings International Conference on Image Processing*, Vol. II, pp. 447–451.

Kundur, D. and Hatzinakos, D. (1996)a Blind image deconvolution, *IEEE Signal Processing Magazine* **13**, 43–64.

Kundur, D. and Hatzinakos, D. (1996)b Blind image deconvolution revisited, *IEEE Signal Processing Magazine* **13**, 61–63.

Lagendijk, R., Biemond, J., and Boekee, D. (1990) Identification and restoration of noisy blurred images using the expectation-maximization algorithm, *IEEE Trans. Acoust. Speech Signal Process.* **38**, 1180–1191.

Molina, R., Vega, M., Abad, J., and Katsaggelos, A. (2003) Parameter estimation in Bayesian high-resolution image reconstruction with multisensors, *IEEE Trans. Image Processing* **12**, 1655–1667.

Myles, Z. and Lobo, N. V. (1998) Recovering affine motion and defocus blur simultaneously, *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**, 652–658.

Nguyen, N., Milanfar, P., and Golub, G. (2001) Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement, *IEEE Trans. Image Processing* **10**, 1299–1308.

Pai, H.-T. and Bovik, A. (2001) On eigenstructure-based direct multichannel blind image restoration, *IEEE Trans. Image Processing* **10**, 1434–1446.

Panci, G., Campisi, P., Colonnese, S., and Scarano, G. (2003) Multichannel blind image deconvolution using the bussgang algorithm: spatial and multiresolution approaches, *IEEE Trans. Image Processing* **12**, 1324–1337.

Park, S., Park, M., and Kang, M. (2003) Super-resolution image reconstruction: a technical overview, *IEEE Signal Proc. Magazine* **20**, 21–36.

Reeves, S. and Mersereau, R. (1992) Blur identification by the method of generalized cross-validation, *IEEE Trans. Image Processing* **1**, 301–311.

Segall, C., Katsaggelos, A., Molina, R., and Mateos, J. (2004) Bayesian resolution enhancement of compressed video, *IEEE Trans. Image Processing* **13**, 898–911.

Shechtman, E., Caspi, Y., and Irani, M. (2005) Space-time super-resolution, *IEEE Trans. Pattern Analysis and Machine Intelligence* **27**, 531–545.

Šroubek, F. and Flusser, J. (2003) Multichannel blind iterative image restoration, *IEEE Trans. Image Processing* **12**, 1094–1106.

Šroubek, F. and Flusser, J. (2005) Multichannel blind deconvolution of spatially misaligned images, *IEEE Trans. Image Processing* **14**, 874–883.

Šroubek, F. and Flusser, J. (2006) Resolution enhancement via probabilistic deconvolution of multiple degraded images, *Pattern Recognition Letters* **27**, 287–293.

Wirawan, Duhamel, P., and Maitre, H. (1999) Multi-channel high resolution blind image restoration, In *Proc. IEEE ICASSP*, pp. 3229–3232.

Woods, N., Galatsanos, N., and Katsaggelos, A. (2003) EM-based simultaneous registration, restoration, and interpolation of super-resolved images, In *Proc. IEEE ICIP*, Vol. 2, pp. 303–306.

Woods, N., Galatsanos, N., and Katsaggelos, A. (2006) Stochastic methods for joint registration, restoration, and interpolation of multiple undersampled images, *IEEE Trans. Image Processing* **15**, 201–213.

Yagle, A. (2003) Blind superresolution from undersampled blurred measurements, In *Advanced*

Signal Processing Algorithms, Architectures, and Implementations XIII, Vol. 5205, Bellingham, pp. 299–309, SPIE.

Zhang, Y., Wen, C., and Zhang, Y. (2000) Estimation of motion parameters from blurred images, *Pattern Recognition Letters* **21**, 425–433.

Zhang, Y., Wen, C., Zhang, Y., and Soh, Y. C. (2002) Determination of blur and affine combined invariants by normalization, *Pattern Recognition* **35**, 211–221.

Zhang, Z. and Blum, R. (2001) A hybrid image registration technique for a digital camera image fusion application, *Information Fusion* **2**, 135–149.

Zitová, B. and Flusser, J. (2003) Image registration methods: a survey, *Image and Vision Computing* **21**, 977–1000.

Zitová, B., Kautsky, J., Peters, G., and Flusser, J. (1999) Robust detection of significant points in multiframe images, *Pattern Recognition Letters* **20**, 199–206.