Combined Superresolution and Blind Deconvolution

Filip Šroubek[∗], Gabriel Cristóbal[†] and Jan Flusser[∗]

[∗]*ÚTIA, Academy of Sciences of CR, Pod vodárenskou věží 4, 182 08 Prague, Czech Republic*
[†]*Instituto de Óptica, CSIC, Serrano 121, 28006 Madrid, Spain*

Abstract. This paper presents a unifying approach to the blind deconvolution and superresolution problem of multiple degraded low-resolution frames of the original scene. We do not assume any prior information about the shape of degradation blurs. The proposed approach consists of building a regularized energy function and minimizing it with respect to the original image and blurs, where regularization is carried out in both the image and blur domains. The image regularization based on variational principles maintains stable performance under severe noise corruption. The blur regularization guarantees consistency of the solution by exploiting differences among the acquired low-resolution images. Experiments on real data illustrate the robustness and utilization of the proposed technique.

Keywords: multichannel systems, blind deconvolution, superresolution, regularized energy minimization
PACS: 42.30.Va

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) $v(x,y)$. Finally, the CCD discretizes the images and produces digitized noisy image $z(x,y)$ (frame). We refer to $z(x,y)$ 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

$$z(x,y) = D[v(x,y) * o(W(x,y))] + n(x,y),\qquad(1)$$

where n is additive noise and W denotes the geometric deformation (warping). $D[\cdot] = S[g * \cdot]$ is the *decimation operator* that models the function of the CCD sensors. It consists of convolution with the *sensor PSF* $g(x,y)$ followed by the *sampling operator S*, which we define as multiplication by a sum of delta functions placed on a evenly spaced grid. The sensor PSF is the PSF of one CCD element. The obtained image $z(x,y)$ can be considered as a discrete image. The above model leads to an image reconstruction problem that is extremely ill-posed. To partially overcome this difficulty, we can assume that multiple LR observations of the original scene are available. Hence we write

$$z_k(x,y) = D[v_k(x,y) * o(W_k(x,y))] + n_k(x,y),\qquad(2)$$

where k is the acquisition index 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 $z_k(x,y)$ are multiple outputs. To our knowledge, this is the most accurate, state-of-the-art model, as it takes all possible degradations into account.

Superresolution (SR) is the process of combining a sequence of LR images in order to produce a higher resolution image or sequence. It is unrealistic to assume that the superresolved image can recover the original scene $o(x,y)$ exactly. A reasonable goal of SR is a discrete version of $o(x,y)$ that has a higher spatial resolution than the resolution of the LR images and that is free of the volatile blurs (deconvolved). In the sequel, we will refer to this superresolved image as a *high resolution (HR) image $u(x,y)$*. The standard SR approach consists of subpixel registration, overlaying the LR images on an HR grid, and interpolating the missing values. The subpixel shift between images thus constitutes the essential assumption. We will demonstrate that introduction of the volatile blurs brings about a more general and robust technique, with the subpixel shift being a special case thereof.

The acquisition model in Eq. (2) embraces three distinct cases frequently encountered in literature. First, we face a registration problem, if we want to resolve the geometric degradation W_k. Second, if the decimation operator D and the geometric transform W_k are not considered, we face a *multichannel* (or multiframe) *blind deconvolution* (MBD) problem. Third, if the volatile blur v_k is not considered or assumed known, and W_k is suppressed up to a subpixel translation, we obtain a classical SR formulation. In practice, it is crucial to consider all three cases at once. We are then confronted with a problem of *blind superresolution* (BSR), which is the subject of this investigation.

Proper registration techniques can suppress large and complex geometric distortions (usually just up to a small between-image shift). There have been hundreds of methods proposed; see e.g. [1] for a survey. So we can assume in the sequel that the LR images are partially registered and that W_k reduces to a small translation.

Current multiframe blind deconvolution techniques, such as [2], 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 since this case violates the standard convolution model. On the contrary, state-of-the-art SR techniques [3] 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 or 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. The 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 that will be derived by the end of Section "Blind Superresolution".

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

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

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 [4]. 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 and then $\mathbf{S}^\varepsilon$ will denote downsampling with the given factor. 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

$$\mathbf{g}_k = \mathbf{DH}_k\mathbf{f} + \mathbf{n}_k = \mathbf{SUH}_k\mathbf{f} + \mathbf{n}_k. \tag{4}$$

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 $\mathbf{C}_F^v\{h_k\}$ and $\mathbf{C}_{F-H+1}^v\{u\}$, 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 [5, 2]. 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.

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 [5, 6], 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 [6]. 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 [5] 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.

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 consists of 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$(\mathcal{N}') = (\hat{H}_1 - H + 1)(\hat{H}_2 - H + 1)$.

The BSR case

We have the sampling matrix $\mathbf{S}^\varepsilon$ that performs downsampling with the factor ε. 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 [7, 8], is to modify (6) in the MBD case by applying downsampling and formulating the problem as

$$\min_{\mathbf{h}} \|\mathcal{N}'[\mathbf{I}_K \otimes \mathbf{S}^\varepsilon \mathbf{U}]\mathbf{h}\|^2, \tag{7}$$

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 (7) 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

$$\mathcal{G} = \mathbf{S}^\varepsilon \mathbf{F} \mathbf{U} \mathcal{H}, \tag{8}$$

where

$$\mathcal{H} := [\mathbf{C}_{\varepsilon E}\{h_1\}(\mathbf{S}^\varepsilon)^T, \ldots, \mathbf{C}_{\varepsilon E}\{h_K\}(\mathbf{S}^\varepsilon)^T], \tag{9}$$

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

The convolution matrix $\mathcal{U}$ has more rows than columns and therefore it is of full column rank (see proof in [5] 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 Null$(\mathcal{G}) \equiv$ 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.,

$$\text{nullity}(\mathcal{G}) \geq KE^2 - (\varepsilon E + H - 1)^2. \tag{10}$$

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

$$\mathbf{N} = \begin{bmatrix} \mathbf{n}_{1,1} & \ldots & \mathbf{n}_{1,N} \\ \vdots & \ddots & \vdots \\ \mathbf{n}_{K,1} & \ldots & \mathbf{n}_{K,N} \end{bmatrix}, \tag{11}$$

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

$$\mathscr{N} := \begin{bmatrix} \mathbf{C}_H\{\tilde{\eta}_{1,1}\} & \ldots & \mathbf{C}_H\{\tilde{\eta}_{K,1}\} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_H\{\tilde{\eta}_{1,N}\} & \ldots & \mathbf{C}_H\{\tilde{\eta}_{K,N}\} \end{bmatrix} \tag{12}$$

and conclude that

$$\mathscr{N}\mathbf{h} = \mathbf{0}, \tag{13}$$

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$(\mathscr{N}) = \varepsilon^4$. For the underestimated blur size, (13) has no solution. For the overestimated blur size $\hat{H}_1 \times \hat{H}_2$, nullity$(\mathscr{N}) = \varepsilon^2(\hat{H}_1 - H + \varepsilon)(\hat{H}_2 - H + \varepsilon)$.

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, the next section will show that $\mathscr{N}$ 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 $\mathscr{G}$ nullity in (10) 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.

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

$$E(\mathbf{f},\mathbf{h}) = \sum_{k=1}^{K} \|\mathbf{D}\mathbf{H}_k\mathbf{f} - \mathbf{g}_k\|^2 + \alpha Q(\mathbf{f}) + \beta R(\mathbf{h}). \tag{14}$$

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 [2] 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

$$Q(\mathbf{f}) = \mathbf{f}^T\mathbf{L}\mathbf{f}, \tag{15}$$

where $\mathbf{L}$ is a kind of high-pass filter. A common strategy is to use convolution with the Laplacian as $\mathbf{L}$, which in the continuous case corresponds to $Q(f) = \int |\nabla f|^2$, or recently, variational integrals $Q(f) = \int \phi(|\nabla f|)$, where ϕ is a strictly convex, nondecreasing function that grows at most linearly.

For the purpose of our discussion it suffices to state that $\mathbf{L}$ 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 (12) contains the correct PSFs h_k in its null space, we define the regularization term as a least-squares fit

$$R(\mathbf{h}) = \|\mathcal{N}\mathbf{h}\|^2 = \mathbf{h}^T \mathcal{N}^T \mathcal{N} \mathbf{h}. \tag{16}$$

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 (14)), 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

$$E(\mathbf{f}, \mathbf{h}) = \sum_{k=1}^{K} \|\mathbf{D}\mathbf{H}_k\mathbf{f} - \mathbf{g}_k\|^2 + \alpha \mathbf{f}^T \mathbf{L}\mathbf{f} + \beta_1 \|\mathcal{N}\mathbf{h}\|^2 + \beta_2 \mathbf{h}^T \mathbf{L}\mathbf{h}. \tag{17}$$

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 (17) 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.

Starting with some initial $\mathbf{h}^0$ the two iterative steps are:

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

$$\Leftrightarrow \quad (\sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{D}^T \mathbf{D}\mathbf{H}_k + \alpha\mathbf{L})\mathbf{f} = \sum_{k=1}^{K} \mathbf{H}_k^T \mathbf{D}^T \mathbf{g}_k, \tag{18}$$

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

$$\Leftrightarrow \quad ([\mathbf{I}_K \otimes \mathbf{F}^T \mathbf{D}^T \mathbf{D}\mathbf{F}] + \beta_1 \mathcal{N}^T \mathcal{N} + \beta_2 \mathbf{L})\mathbf{h} = [\mathbf{I}_K \otimes \mathbf{F}^T \mathbf{D}^T]\mathbf{g}, \tag{19}$$

where $\mathbf{F} := \mathbf{C}_H^v\{f\}$, $\mathbf{g} := [\mathbf{g}_1^T, \ldots, \mathbf{g}_K^T]^T$ and m is the iteration step. Note that both steps consist of simple linear equations. To improve convergence we add feasible regions for the HR image and PSFs specified as lower and upper bounds constraints.

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 [10] or expectation maximization [11], 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.

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 Gaussian 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.

Simulated data

First, let us demonstrate the BSR performance with a simple experiment. A 150×230 image blurred with the six masks (see Fig. 1.a) 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 (see one example in Fig. 1.b) as an input, we estimated the original HR image with the proposed BSR algorithm for $\varepsilon = 1.25$ (Fig. 1.c) and 1.75 (Fig. 1.d). The HR image for $\varepsilon = 1.25$ 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 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 [2] followed by 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 [12, 13]. This method uses a MAP framework for the joint estimation of image registration parameters

a. original image   b. LR   c. $\varepsilon = 1.25$ d. $\varepsilon = 1.75$
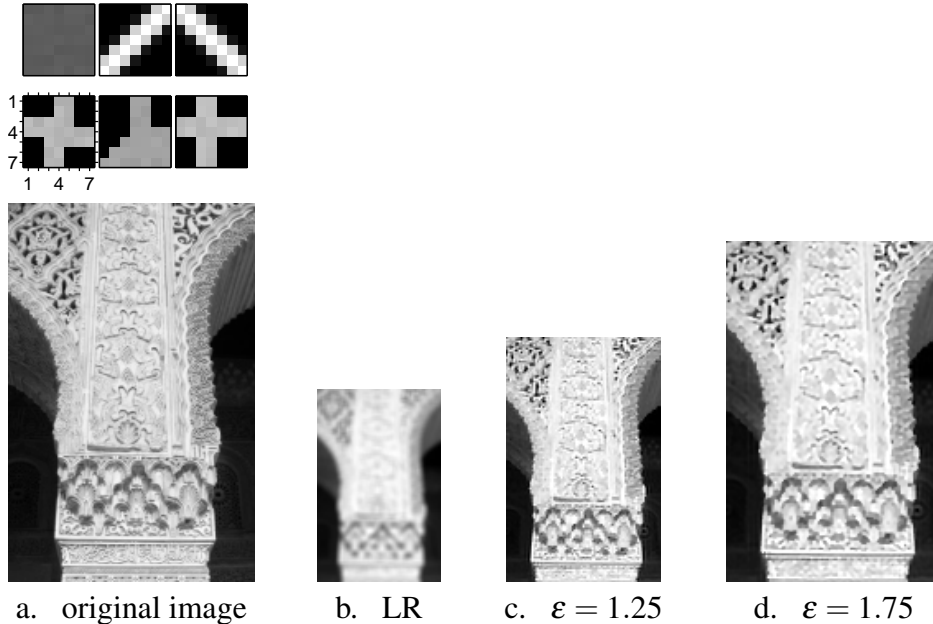
FIGURE 1. Simulated data: (a) original 150×230 image and six 7×7 volatile PSFs used to blur the image; (b) one of six LR images with the downsampling factor 2; (bc BSR for $\varepsilon = 1.25$; (d) BSR for $\varepsilon = 1.75$.

(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 [4].

In the case of BSR, we have seen that two distinct approaches exist for blur estimation. Either we use the naive approach in (7) that directly utilizes the MBD formulation, or we apply the intrinsically SR approach summarized in (13). 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. 1.a, six LR images 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. 2. The standard SR approach in Fig. 2.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. 2.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 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

---

[1] The signal-to-noise ratio is defined as SNR $= 10\log(\sigma_f^2/\sigma_n^2)$, where σ_f and σ_n are the image and noise standard deviations, respectively.
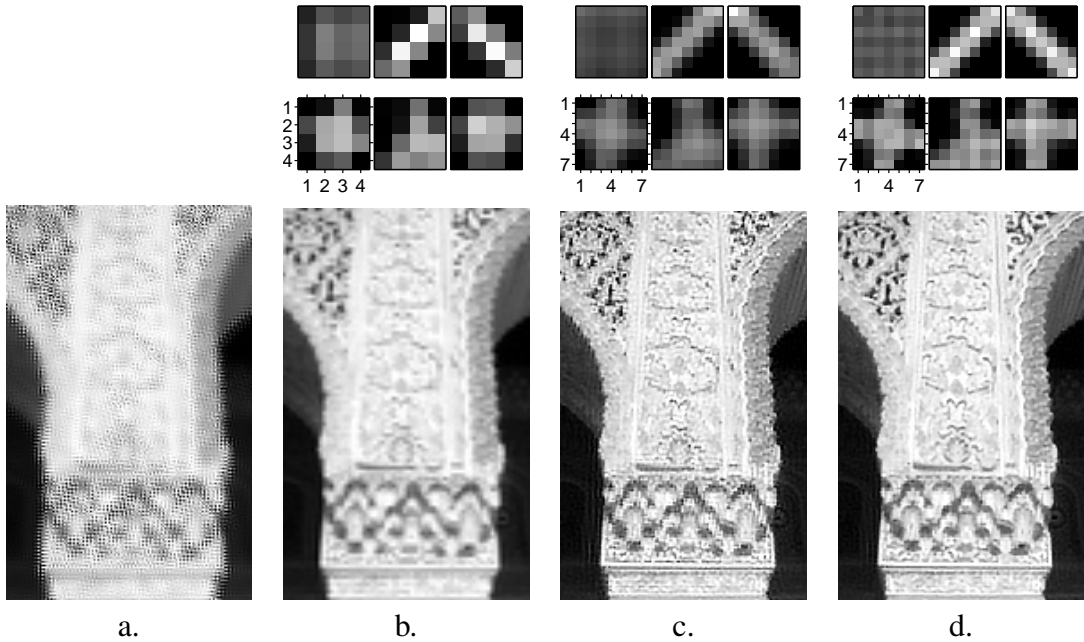
FIGURE 2. Comparison of four different SR approaches ($\varepsilon = 2$): (a) standard SR method, (b) MBD followed by bilinear interpolation, (c) naive BSR approach and (d) proposed intrinsic BSR approach. Volatile blurs estimated by each method, except in the case of standard SR, are in the top row. 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.

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. 2.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. 2.d.

Real data

The next two experiments demonstrate the true power of our BSR 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 $1/10$s. The digital camera was 5 Mpixel Olympus C5050Z equipped with $3\times$ 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. 3.a. The other frames look similar. The long shutter speed ($1/10$s) 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. 3.b. Note that many facial features (eyes, glasses, mouth) indistinguishable in the original LR image became visible in the HR image.

a.   b.

FIGURE 3. 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. One of the LR images enlarged with zero-order interpolation is in Fig. 4.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 $1.7\times$ (close to the desired SR factor $5/3$). This image served as the ground truth; see Fig. 4.c. The proposed BSR method returned a well reconstructed HR image (Fig. 4.b), which is comparable to the ground truth acquired with the optical zoom.

CONCLUSIONS

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. The last two steps were the subject of our investigation. We proposed a unifying system that simultaneously estimates image blurs and recovers the original undistorted image, all in high resolution, without any

a.   b.   c.

FIGURE 4. Reconstruction of images acquired with a digital camera ($\varepsilon = 5/3$): (a) one of eight LR images enlarged by zero-order interpolation; (b) HR image estimated by the BSR algorithm; (c) image acquired with optical zoom $1.7\times$. The BSR algorithm achieved reconstruction comparable to the image with optical zoom.

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.

ACKNOWLEDGMENTS

REFERENCES

1. B. Zitová, and J. Flusser, *Image and Vision Computing* **21**, 977–1000 (2003).
2. F. Šroubek, and J. Flusser, *IEEE Trans. Image Processing* **14**, 874–883 (2005).
3. S. Park, M. Park, and M. Kang, *IEEE Signal Proc. Magazine* **20**, 21–36 (2003).
4. D. Capel, *Image Mosaicing and Super-Resolution*, Springer, New York, 2004.
5. G. Harikumar, and Y. Bresler, *IEEE Trans. Image Processing* **8**, 202–219 (1999).
6. G. Giannakis, and R. Heath, *IEEE Trans. Image Processing* **9**, 1877–1896 (2000).
7. F. Šroubek, and J. Flusser, *Pattern Recognition Letters* **27**, 287–293 (2006).
8. Y. Chen, Y. Luo, and D. Hu, "A general approach to blind image super-resolution using a PDE framework," in *Proc. SPIE*, 2005, vol. 5960, pp. 1819–1830.
9. G. Aubert, and P. Kornprobst, *Mathematical Problems in Image Processing*, Springer Verlag, New York, 2002.
10. N. Nguyen, P. Milanfar, and G. Golub, *IEEE Trans. Image Processing* **10**, 1299–1308 (2001).
11. R. Molina, M. Vega, J. Abad, and A. Katsaggelos, *IEEE Trans. Image Processing* **12**, 1655–1667 (2003).
12. R. Hardie, K. Barnard, and E. Armstrong, *IEEE Trans. Image Processing* **6**, 1621–1633 (1997).
13. C. Segall, A. Katsaggelos, R. Molina, and J. Mateos, *IEEE Trans. Image Processing* **13**, 898–911 (2004).