Image Fusion via Probabilistic Deconvolution

F. Sroubek and J. Flusser
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic
e-mail: {sroubekf,flusser}@utia.cas.cz

ABSTRACT

We present a maximum a posteriori solution to the problem of obtaining a high-resolution image from a set of degraded low-resolution images of the same scene. The proposed algorithm has the advantage that no prior knowledge of blurring functions is required and it can handle unknown misregistrations between the input images. An efficient implementation scheme of alternating minimizations is presented together with experiments that demonstrates the performance of the algorithm.

I. INTRODUCTION

Image fusion is one of quickly developing advanced methods used in processing of remotely sensed images. The term *fusion* means in general an approach to extraction of information spontaneously adopted in several domains. The goal of image fusion is to integrate complementary multisensor, multitemporal and/or multiview information into one new image containing information the quality of which cannot be achieved otherwise. The term "quality" depends on the application requirements.

In remote sensing, image fusion has been mainly used to achieve high spatial and spectral resolutions by combining images from two sensors, one of which has high spatial resolution and the other one high spectral resolution. Typically, a multispectral satellite image (like SPOT XMS for instance) is fused with a high-resolution panchromatic image (SPOT Panchro) or with an aerial image. Most fusion methods are based on image decomposition into low-pass and high-pass bands and on combining the low-pass band of the multispectral image with the high-pass band(s) of the panchromatic image [1]–[5]. Similar effect can be achieved by transforming the multispectral image into IHS coordinates and replacing intensity component by the panchromatic image [6], [7].

In this paper, we consider a different problem formulation. Assuming two or more low-resolution images from the same sensor (or from different sensors of the same type) are available, our goal is to obtain fused image of higher spatial resolution than the resolution of the input channels. Contrary to the previous case, this task is more complicated because we do not have the high-resolution information in any form. This problem appears in remote sensing very often. Due to the physical limitations of the sensors (see, for instance, [8] for detailed explanation) and imperfect observational conditions, the acquired images represent only a degraded versions of the original scene, where mainly the high-frequency information is suppressed, degraded or missing. Fusion of the low-resolution images together is an effective means of breaking the sensor limits and of removing the degradation introduced by atmospheric turbulence, sensor motion, and other factors.

We should point out that this problem appears also outside the area of remote sensing (it has attracted the attention of producers of low-resolution cameras and videos, among others) and has lead to developing techniques known as *multichannel blind deconvolution* and *superresolution imaging*, see [9]–[16] for a survey and other references. However, a common weakness of the previous techniques is that they need too much *a priori* information which is not realistic in practice. For instance, they require a knowledge of a shape and size of the blurring function, availability of a high-resolution reference frame, or accurate geometric alignment (registration) of the input channels.

In this paper we present a fusion method based on a stochastic approach. The optimal solution is defined as a maximum *a posteriori* (MAP) estimate and is found by an alternating minimization (AM) algorithm. The main feature of the new fusion method is that it does not require any knowledge of the blurring functions and the input channels might be mutually shifted by an unknown vector. Allowing only translational between-channel misregistration is not a serious limitation. Larger and more complex geometric distortions can be suppressed (usually just up to a small between-image shift) by a proper registration method (there have been hundreds of them investigated, see [17] for a survey).

II. MAP ANALYSIS

Let us assume that the k-th acquired low-resolution image (channel) z_k can be modelled by blurring the "ideal" image u and shifting the result by an unknown vector $(a_k, b_k) = t_k$, i.e.,

$$z_k(x + a_k, y + b_k) = (u * h_k)(x, y) + n_k(x, y), \qquad (1)$$

where h_k is an unknown PSF having a character of a low-pass filter, and n_k denotes additive noise. This model is a very realistic description of many low-resolution satellite sensors [8]. In the discrete domain, this degradation model takes the form

$$\mathbf{z}_k = \mathbf{T}_k \mathbf{H}_k \mathbf{u} + \mathbf{n}_k\,, \;\; k = 1, \ldots, K,$$

where $\mathbf{z}_k$, $\mathbf{u}$, and $\mathbf{n}_k$ are discrete lexicographically ordered equivalents of image functions z_k, u, and n_k, respectively.

$\mathbf{T}_k$ is a translation operator shifting an image by t_k pixels, i.e. a linear filter with the delta function at the position t_k. One can readily see that the matrix product $\mathbf{T}_k\mathbf{H}_k = \mathbf{G}_k$ defines convolution with a mask $\mathbf{g}_k$ that is a shifted version of a mask $\mathbf{h}_k$ (discrete representation of h_k). By concatenating the channels, the previous equation can be rewritten in two equivalent forms

$$\mathbf{z} = \mathbf{G}\mathbf{u} + \mathbf{n} = \mathbf{U}\mathbf{g} + \mathbf{n}\,, \tag{2}$$

where $\mathbf{z} \equiv [\mathbf{z}_1^T, \dots \mathbf{z}_K^T]^T$, $\mathbf{G} \equiv [\mathbf{G}_1^T, \dots, \mathbf{G}_K^T]^T$, $\mathbf{n} \equiv [\mathbf{n}_1^T, \dots, \mathbf{n}_K^T]^T$, $\mathbf{g} \equiv [\mathbf{g}_1^T, \dots, \mathbf{g}_K^T]^T$, and $\mathbf{U}$ is a block-diagonal matrix with K blocks each performing convolution with the image $\mathbf{u}$.

Adopting a stochastic approach, the problem of image fusion can be formulated as an MAP estimation. We assume that the images $\mathbf{u}$, $\mathbf{g}$ and $\mathbf{z}$ are random vector fields with given probability density functions (pdf) $p(\mathbf{u})$, $p(\mathbf{g})$ and $p(\mathbf{z})$, respectively, and we look for such realizations of $\mathbf{u}$ and $\mathbf{g}$ which maximize the *a posteriori* probability $p(\mathbf{u}, \mathbf{g}|\mathbf{z})$. According to the Bayes rule, the relation between *a priori* probabilities $p(\mathbf{u})$, $p(\mathbf{g})$ and the a posteriori probability is $p(\mathbf{u}, \mathbf{g}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{u}, \mathbf{g})p(\mathbf{u})p(\mathbf{g})$. The conditional pdf $p(\mathbf{z}|\mathbf{u}, \mathbf{g})$ follows from (2) and from our assumption of white Gaussian noise, i.e.

$$p(\mathbf{z}|\mathbf{u}, \mathbf{g}) \propto \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{G}\mathbf{u})^T\Sigma^{-1}(\mathbf{z} - \mathbf{G}\mathbf{u})\right\}\,,$$

where Σ is the noise diagonal covariance matrix with $\{\sigma_k^2\}_{k=1}^K$ on the corresponding positions on the main diagonal. If the same noise variance σ^2 is assumed in each channel, Σ^{-1} reduces to a scalar σ^{-2}.

A. A priori *distribution of the original image*

A general model for the prior distribution $p(u)$ is a Markov random field which is characterized by its Gibbs distribution given by $p(u) \propto \exp(-F(u)/\lambda)$, where λ is a constant and F is called the *energy function*. One can find various forms of the energy function in the literature, however, the most promising results have been achieved for variational integrals. The energy function then takes the form

$$F(u) = \int \phi(|\nabla u|)\,, \tag{3}$$

where ϕ is 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) or $\log(\cosh(s))$. Nonconvex functions may behave an unpredictable manner but since they often provide better results for segmentation problems, forms, such as $\log(1+s^2)$, $s^2/(1+s^2)$ or $\arctan(s^2)$ (Mumford-Shah functional), are often used as well. The energy function based on the variational integral is highly nonlinear and to overcome this difficulty we follow a half-quadratic scheme described in [18] which introduces an auxiliary variable. A special attention must be paid to the discretization of the image gradient ∇u and relaxation of ϕ. In addition, we confine the distribution to an amplitude constraint set $C_u \equiv \{u|\alpha \le u \le \beta\}$ with amplitude bounds derived from the input images, typically $\alpha = 0$ and $\beta = 255$. We thus define the prior distribution as

$$p(\mathbf{u}) = \begin{cases} \frac{1}{Z}\exp\left\{-\frac{1}{2\sigma_u^2}\mathbf{u}^T\mathbf{L}(v)\mathbf{u}\right\} & \text{if } \mathbf{u} \in C_u\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

where Z is the partition function, σ_u^2 denotes the image variance, $\mathbf{u}^T\mathbf{L}(v)\mathbf{u}$ represents the discretization of (3) and v is the auxiliary variable introduced by the half-quadratic scheme, which is calculated as

$$v(x, y) = \frac{\phi'(|\nabla u(x, y)|)}{|\nabla u(x, y)|}\,. \tag{4}$$

Matrix $\mathbf{L}(v)$ is a positive semidefinite block tridiagonal matrix constructed by v that performs shift-variant convolution with v.

B. A priori *distribution of the blurs*

The shape of the prior distribution $p(\mathbf{g})$ can be derived from the fundamental multichannel constraint stated in [9], [10]. Let $\mathbf{Z}_k$ denote the convolution matrix with the degraded image $\mathbf{z}_k$. If noise $\mathbf{n}_k$ is zero and the original channel masks $\{\mathbf{h}_k\}$ are *weakly coprime*, i.e. their only common factor is a scalar, then all solutions $\{\hat{\mathbf{g}}_k\}$ to

$$\mathbf{Z}_i\hat{\mathbf{g}}_j - \mathbf{Z}_j\hat{\mathbf{g}}_i = \mathbf{0}\,, \quad 1 \le i < j \le K \tag{5}$$

have the following forms. Let S_g denote the sum of the maximum blur size and the maximum shift between the channels. If S_g is known the solution equals $\{\alpha\mathbf{g}_k\}$ for any scalar α. If S_g is not known, it must be first estimated and two distinct situations arise. If S_g is underestimated, zero vector is the only solution of (5). If S_g is overestimated, then the space of all solutions of (5) contains the correct masks $\{\mathbf{g}_k\}$ and the dimensionality of this solution space is proportional to the degree of the overestimation. After further stacking the system of equations (5), we obtain

$$\mathcal{Z}\hat{\mathbf{g}} = \mathbf{0}\,, \tag{6}$$

where $\hat{\mathbf{g}} \equiv [\hat{\mathbf{g}}_1^T, \dots, \hat{\mathbf{g}}_K^T]^T$. If the noise term $\mathbf{n}_k$ is present, it follows from (2) that the left-hand side of (6) equals a realization of a Gaussian process of zero mean and covariance $\mathcal{C} = \mathcal{G}\Sigma\mathcal{G}^T$, where $\mathcal{G}$ takes the form of $\mathcal{Z}$ in (6) with $\mathbf{Z}_i$ replaced by $\mathbf{G}_i$.

It is desirable to include also other prior knowledge about the blurs, such as positivity or constant energy. We therefore define a set of admissible solutions as $C_g \equiv \{\mathbf{g}|g_k(x, y) \ge 0 \wedge \sum_{x,y} g_k(x, y) = 1,\ k = 1, \dots, K\}$ and propose the following prior distribution:

$$p(\mathbf{g}) = \begin{cases} \frac{1}{Z}\exp\left\{-\frac{1}{2}\mathbf{g}^T\mathcal{Z}^T\mathcal{C}^{-1}\mathcal{Z}\mathbf{g}\right\} & \text{if } \mathbf{g} \in C_g\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

The inverse of the matrix $\mathcal{C}$ is not trivial and the matrix is constructed by the blurs $\mathbf{g}$ that are to be estimated. To overcome this difficulty, we approximate $\mathcal{C}$ by a diagonal matrix $\mathcal{D}$ such that $\mathrm{diag}(\mathcal{D}) = \mathrm{diag}(\mathcal{C})$, where $\mathrm{diag}(\cdot)$ denotes

the main diagonal of the matrix. The elements of $\mathcal{D}$ take the form $\sigma_i^2\|\mathbf{g}_j\|^2 + \sigma_j^2\|\mathbf{g}_i\|^2$ for $1 \le i < j \le K$. The value of $\|\mathbf{g}_i\|^2$ is not known in advance, but a good initial approximation can be given. Since $\mathbf{g} \in C_g$, $\frac{1}{\prod S_g} \le \|\mathbf{g}_i\|^2 \le 1$ and we use the bottom limit for $\|\mathbf{g}_i\|^2$.

C. AM-MAP algorithm

The MAP estimation is given by

$$\{\hat{\mathbf{u}}, \hat{\mathbf{g}}\} = \arg \min_{\mathbf{u}\in C_u, \mathbf{g}\in C_g} \left\{ (\mathbf{z} - \mathbf{G}\mathbf{u})^T \Sigma^{-1} (\mathbf{z} - \mathbf{G}\mathbf{u}) + \frac{1}{\sigma_u^2} \mathbf{u}^T \mathbf{L}(v)\mathbf{u} + \mathbf{g}^T \mathcal{Z}^T \mathcal{D}^{-1} \mathcal{Z}\mathbf{g} \right\}. \quad (7)$$

Such problems can be solved by means of genetic algorithms, e.g. simulated annealing. In this paper we have adopted an approach of alternating minimizations over $\mathbf{u}$ and $\mathbf{g}$. The advantage of this scheme lies in its simplicity. Each term in (7) is convex and the derivatives w.r.t. $\mathbf{u}$ and $\mathbf{g}$ can be easily calculated. The proposed AM–MAP algorithm alternates between two steps

1. $\mathbf{u} = (\mathbf{G}^T\Sigma^{-1}\mathbf{G} + \frac{1}{\sigma_u^2}\mathbf{L}(v))^{-1}\mathbf{G}^T\Sigma^{-1}\mathbf{z} \wedge \mathbf{u} \in C_u,$
2. $\mathbf{g} = (\mathbf{U}^T\Sigma^{-1}\mathbf{U} + \mathcal{Z}^T\mathcal{D}^{-1}\mathcal{Z})^{-1}\mathbf{U}^T\Sigma^{-1}\mathbf{z} \wedge \mathbf{g} \in C_g\,.$

In step 1 the flux variable v is updated according to (4). Inversion of the matrix in step 1 cannot be carried out directly because of the matrix size. Instead, we use the method of conjugate gradients to find the solution. Both $\mathbf{U}^T\Sigma^{-1}\mathbf{U}$ and $\mathcal{Z}^T\mathcal{D}^{-1}\mathcal{Z}$ have the size squared proportional to the blur size and can be constructed directly without building the individual matrices $\mathbf{U}$ and $\mathcal{Z}$ that have much larger size.

One must supply the blur size to the algorithm. An important feature is that an accurate estimation is not necessary; we must only guarantee that the blur size is not underestimated. In addition, the noise covariance Σ and the image variance σ_u^2 are obligatory in the algorithm. However, if noise has the same variance σ_n^2 in every channel, the MAP expression (7) is simplified and only the signal to noise ratio σ_u^2/σ_n^2 is required.

III. Experimental results

Experiments were carried out on two simulated scenarios, seemingly acquired at different time instances. For instance, one can imagine they were captured by a low-resolution sensor similar to AVHRR.

As the input for the first simulation (playing the role of the "ideal" image), the 300×300 SPOT HRV image covering the north-western part of Prague (Czech capital) was used (see Fig. 1(c)). We chose this image as a typical representation of urban areas. To simulate two low-resolution acquisitions, the image was blurred by two randomly generated 5×5 motion masks, corrupted by Gaussian additive noise of SNR = 50 dB, and mutually shifted by 5 pixels in both directions (see Figs. 1(a) and 1(b)). The AM-MAP algorithm was initialized with the overestimated blur size 15×15. Fig. 1(d) depicts the fused image.

The second experiment was conducted in a similar way but on a image showing agricultural areas; see Fig. 2(c). The parameters of the simulation were the same except this time the AM-MAP algorithm was initialized with the less overestimated blur size 12×12. The fused images is shown in Fig. 2(d).

The fused images are by visual comparison much sharper than the input channels and are fully comparable to the "ideal" images, which demonstrates excellent performance of the method. The degree of overestimation does not have a serious impact on the quality of the fused image.

IV. Conclusions

We have developed an iterative fusion algorithm that recovers a high-resolution image from misaligned and blurred input channels. The fusion problem is formulated as the MAP estimation with the prior probabilities derived from the variational integral and from the mutual relation of coprime channels. The presented experiments indicate that this approach provides high-quality fused images, fully comparable to the "ideal" ones. We envisage that the possible future extension of the proposed method is to include resolution enhancement of the fused image, i.e. solve the super-resolution and blind deconvolution problems simultaneously.

Acknowledgments

Financial support of this research was provided by the Grant Agency of the Czech Republic under the project No. 102/04/0155.

References

[1] B. Duport, J. Girel, J. Chassery, and G. Pautou, "The use of multiresolution analysis and wavelets transform for merging SPOT panchromatic and multispectral image data," *Photogrammetric Engineering & Remote Sensing*, vol. 69, no. 9, pp. 1057–1066, Sept. 1996.

[2] J. Núñez, X. Otazu, and O. Fors, "Image fusion with additive multiresolution wavelet decomposition. Applications to SPOT+Landsat images," *J. Opt. Soc. Am. A*, vol. 16, no. 3, pp. 467–474, Mar. 1999.

[3] P. Scheunders and S. D. Backer, "Fusion and merging of multispectral images with use of multiscale fundamental forms," *J. Opt. Soc. Am. A*, vol. 18, no. 10, pp. 2468–2477, Oct. 2001.

[4] S. Li, J. Kwok, and Y. Wang, "Using the discrete wavelet frame transform to merge Landsat TM and SPOT panchromatic images," *Information Fusion*, vol. 3, pp. 17–23, 2002.

[5] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion – the ARSIS concept and some successful implementation schemes," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 58, pp. 4–18, 2003.

[6] J. Carper, T. Lillesand, and R. Kiefer, "The use of Intensity-Hue-Saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.

[7] P. Chavez, S. Sides, and J. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic," *Photogrammetric Engineering & Remote Sensing*, vol. 57, pp. 295–303, 1991.

[8] S. E. Reichenbach, D. E. Koehler, and D. W. Strelow, "Restoration and reconstruction of AVHRR images," *IEEE Trans. Geoscience and Remote Sensing*, vol. 33, pp. 997–1007, 1995.

[9] G. Harikumar and Y. Bresler, "Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms," *IEEE Trans. Image Processing*, vol. 8, no. 2, pp. 202–219, Feb. 1999.

[10] G. Giannakis and R. Heath, "Blind identification of multichannel FIR blurs and perfect image restoration," *IEEE Trans. Image Processing*, vol. 9, no. 11, pp. 1877–1896, Nov. 2000.

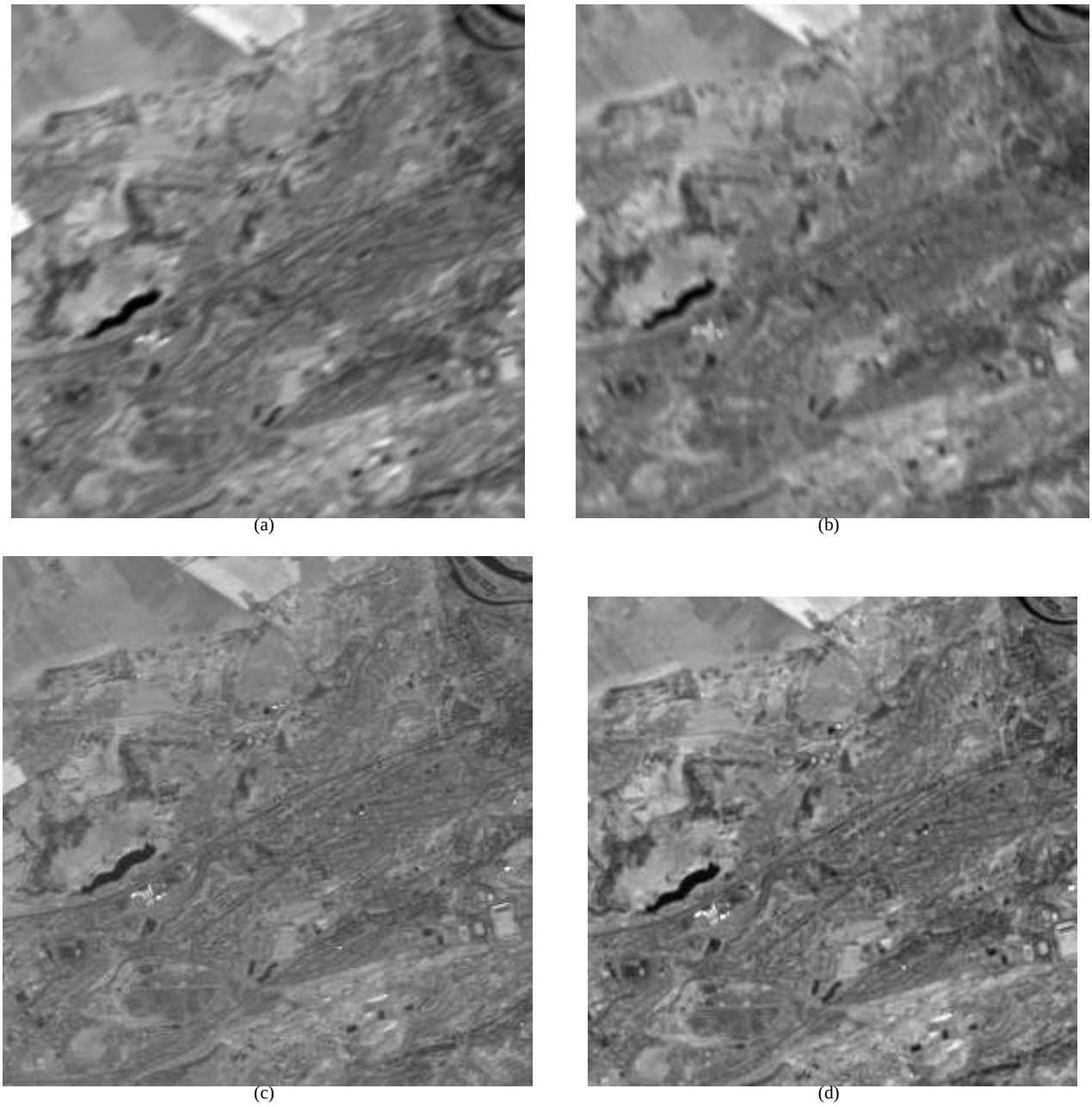

Fig. 1. AM-MAP fusion of urban-area images: (a)-(b) simulated low-resolution images; (c) "ideal" image for comparison; (d) result of fusing (a) and (b). The fused image is of smaller size since only the overlap areas of the input images can be fused.

[11] H.-T. Pai and A. Bovik, "On eigenstructure-based direct multichannel blind image restoration," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1434–1446, Oct. 2001.

[12] A. Rav-Acha and S. Peleg, "Restoration of multiple images with motion blur in different directions," in *IEEE Workshop on Applications of Computer Vision (WACV)*, 2000, pp. 22–27.

[13] G. Panci, P. Campisi, S. Colonnese, and G. Scarano, "Multichannel blind image deconvolution using the bussgang algorithm: Spatial and multiresolution approaches," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1324–1337, Nov. 2003.

[14] F. Šroubek and J. Flusser, "Multichannel blind iterative image restoration," *IEEE Trans. Image Processing*, vol. 12, no. 9, pp. 1094–1106, Sept. 2003.

[15] S. Park, M. Park, and M. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Proc. Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[16] N. Woods, N. Galatsanos, and A. Katsaggelos, "EM-based simultaneous registration, restoration, and interpolation of super-resolved images," in *Image Processing, 2003. Proceedings.*, vol. 2, 2003, pp. 303–306.

[17] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[18] G. Aubert and P. Kornprobst, *Mathematical Problems in Image Processing*. New York: Springer Verlag, 2002.

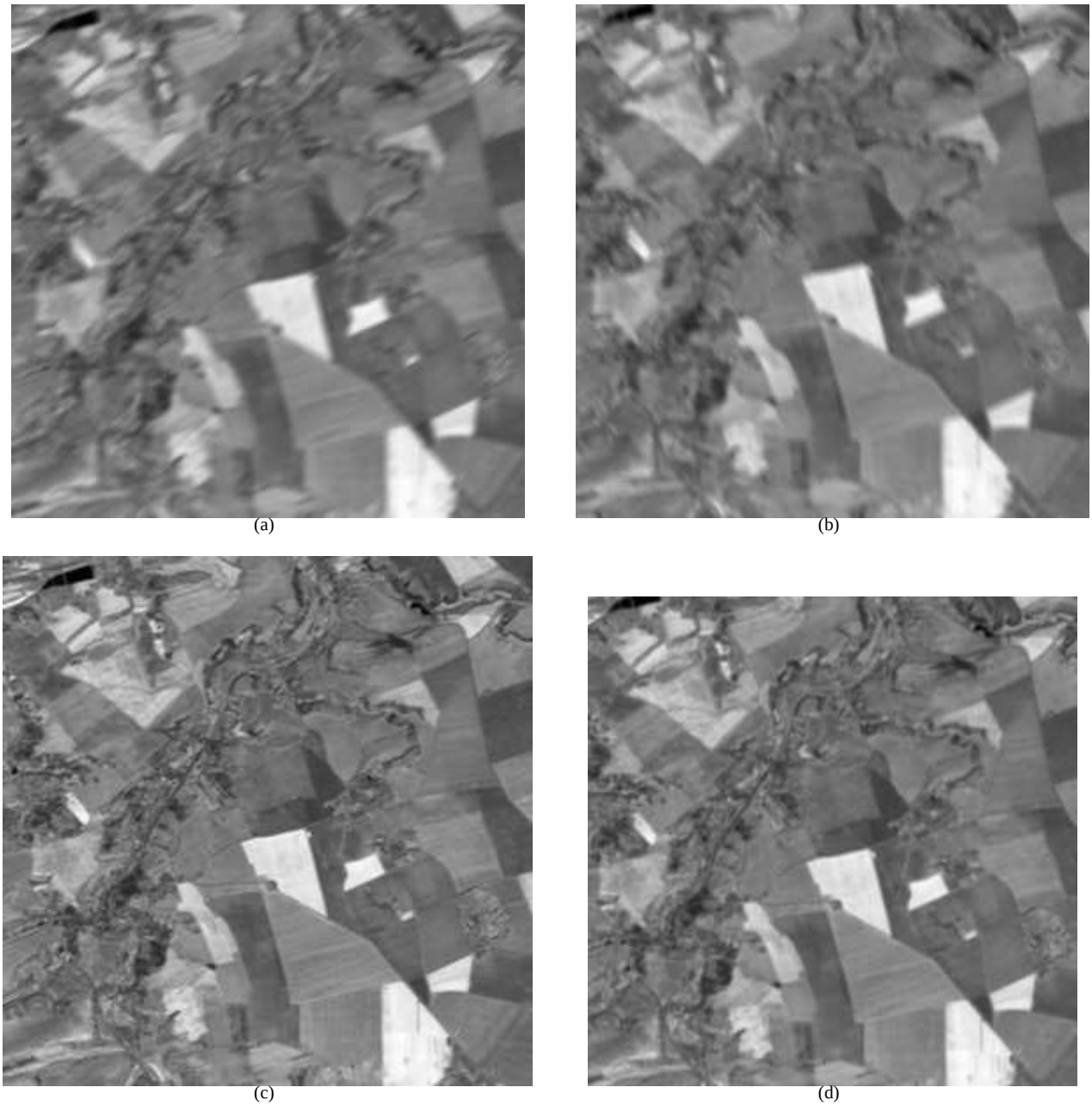

Fig. 2. AM-MAP fusion of agricultural-area images: (a)-(b) simulated low-resolution images; (c) ”ideal” image for comparison; (d) result of fusing (a) and (b). The fused image is of smaller size since only the overlap areas of the input images can be fused.