Multifocus Fusion with Oriented Windows

F. Sroubek[a], S. Gabarda[b], R. Redondo[b], S. Fischer[b] and G. Cristóbal[b]

[a]Academy of Sciences, Pod vodárenskou věží 4, Prague, Czech Republic;
[b]Instituto de Óptica, CSIC, Serrano 121, 28006 Madrid, Spain

ABSTRACT

A wide variety of image fusion techniques exist. A key term that is common to most is the "*decision map*". This map determines which information to take and at what place. Multifocus fusion deals with a stack of images that were acquired with a different focus point. In this case, one can say that the task of the decision map is to label parts that are in focus. If the focus length for each image in the stack is known, the decision map determines also a depth map that can be used for 3D surface reconstruction. Accuracy of the decision map is critical not only for image fusion itself, but even more for the surface reconstruction. Erroneous decisions can produce unrealistic glitches. We propose here to use information about image edges for increasing the accuracy of the decision map and enhancing in this way a standard wavelet-based fusion approach. We demonstrate the performance on real multifocus data under different noise levels.

Keywords: multifocus image fusion, wavelets

1. INTRODUCTION

The term *fusion* means in general an approach to extract information spontaneously from several sources. The goal of image fusion is to integrate complementary multisensor, multitemporal and/or multiview data into a new image containing information, the quality of which cannot be achieved otherwise. The term "quality" depends on the application requirements. The individual images entering the fusion process are called *channels*. Image fusion has been used in many application areas, e.g., in remote sensing and astronomy, in machine vision and mobile robot navigation, in automatic change detection and monitoring of dynamic processes, and last but not least in optical microscopy (multifocus fusion) and medical imaging (multimodal fusion).

Image fusion usually starts with dividing the channels into subregions, calculating a measure of information level in the regions (in the literature often referred to as a *activity level*) (AL) and then utilizing some fusion rules to combine the channels. The channel comparison can be done at different levels of abstraction.[1] The lowest possible is the pixel level, which refers to the merging of measured physical parameters (intensity values of pixels). One step higher is feature-level fusion, which operates on characteristics such as size, shape, edge, contrast and texture. The highest level of abstraction, called decision level fusion, deals with symbolic representations of images. When we talk about image fusion we usually refer to fusion that lies between the pixel and feature level. A common apparatus in image fusion is a multiscale transform (MST), such as the Laplacian pyramid, contrast pyramid, gradient pyramid and wavelet decomposition. Coefficients of MST can be regarded as simple features. The measure of information level in the subregion is the crucial point in the whole process and several different methods were suggested in the literature. In most of the cases, the AL is proportional to the energy of high frequencies in the channel. It corresponds with an intuitive expectation that high frequencies contain details that are important for our visual perception and understanding of the fused image. Image variance, norm of image gradient, norm of image Laplacian,[2] energy of a Fourier spectrum,[3] image moments,[4] and energy of high-pass bands of a wavelet transform[5–7] belong to the most popular measures of AL. Another important issue is the way to divide the channels into subregions. The simplest but the most common strategy is to use square neighborhoods around each image position. More advanced approaches propose to perform first segmentation of the channels and then use the obtained segments as subregions. At each subregion (or pixel neighborhood),

Further author information: (Send correspondence to G. Cristóbal)
G. Cristóbal: E-mail: gabriel@optica.csic.es
F. Sroubek is currently with the Institute de Óptica, CSIC, Madrid, Spain; E-mail: filip@optica.csic.es

AL's of all channels are compared and the information (pixel values or MST coefficients) of the channel with the highest activity is preserved (*maximum selection rule*). By this process we create the decision map (DM). Alternatively, the first couple of channels with the highest activity can be preserved and their information is averaged. A consistency verification stage follows to prevent occurrence of outlying decisions. One can regard this step as smoothening of the DM. Once the DM is ready, we create the multiscale representation of the fused image and perform the inverse MST. An excellent overview of multiscale image fusion is given in Ref.[8]

This paper concerns multifocus fusion, i.e., we fuse images that depict the same scene but each image was acquired with a different focus length. In this case, AL is often referred to *focus measure* and DM identifies regions in focus. One must assume that there exists a partitioning of the scene into regions and each region is acquired undistorted (in focus) in at least one channel. The identification of undistorted subregions determines the distance of the subregions from camera's (or microscope's) objective lens. Then the distance can be used for surface reconstruction of the measured object (*2.5D reconstruction*). An accurate DM is not only important for valid reconstruction of the fused image, but it is also critical for the surface reconstruction. Erroneous decisions can produce unrealistic peaks and valleys on the surface. A way to increase the accuracy of DM in multifocus fusion is to integrate into the calculation of focus measure some additional information about the characteristics of the scene in question. One such possibility is to use information about edges in the input images. Borders of objects in the scene often delineate abrupt changes in the scene depth and therefore correspond to changes in DM. Erroneous decisions in DM tend to occur along the changes of the scene depth. The main source of these errors is due to the fact that neighborhoods, on which we calculate the focus measure, are fixed and thus they can cover regions of different depth. To eliminate this problem, we have proposed to use adjustable (or oriented) window neighborhoods that are elongated in the direction parallel to the edge in order to minimize the probability that the neighborhood will cross into another region. We thus move one step further towards feature-level fusion.

In the next section we review the image fusion techniques that use MST's. Section 3 introduces the concept of oriented windows and incorporates it in the wavelet-based fusion procedure. In Section 4 experiments on real data are presented and comparison with a standard wavelet-based fusion is given.

2. MULTISCALE-BASED FUSION

First we give a brief description of the fusion methodology based on the multiscale decompositions. More or less we follow the notation and terminology given in Ref.[6] Let $\mathbf{I}_j$ denote the j-th input channel and $\mathbf{Z}$ denote the fused image. The coefficients of MST can be addressed with a multi-index $\vec{\mathbf{p}} = (m, n, k, l)$, where m, n indicate the spatial position in a given frequency band k and l the decomposition level. In the case of the standard wavelet transform $k = 1, 2, 3$ except the last level, where we have only one low-pass band $W_j(m, n, 1, l_{\max})$ (approximation of the signal). We denote the MST transform of $\mathbf{I}_j$ as $W_j(\vec{\mathbf{p}})$. The activity level and the decision map have the same structure as W_j and we denote them as $A_j(\vec{\mathbf{p}})$ and $D(\vec{\mathbf{p}})$ respectively. The complete procedure of multiscale-based fusion is depicted in Fig. 1. In this figure fusion of two input channels with one-level wavelet decomposition is assumed. However, the same scheme applies for any MST and any number of input channels.

The AL of an MST coefficient reflects the local energy in the area spanned by this coefficient in the original image. The AL is related to the absolute or squared value of the corresponding coefficients in the MST domain. The simplest form is to consider each coefficient separately, i.e.,

$$A_j(\vec{\mathbf{p}}) = |W_j(\vec{\mathbf{p}})| \,. \tag{1}$$

This is however not robust against noise and therefore window-based ALs were introduced that employ a small (typically 3x3 or 5x5) window centered at the current coefficient position. This approach can be further generalized and leads to weighted averages:

$$A_j(\vec{\mathbf{p}}) = \sum_{s,t} w(s,t) |W_j(m+s, n+t, k, l)| \,, \tag{2}$$

where $w(s,t)$ is a weight and $\sum_{s,t} w(s,t) = 1$.

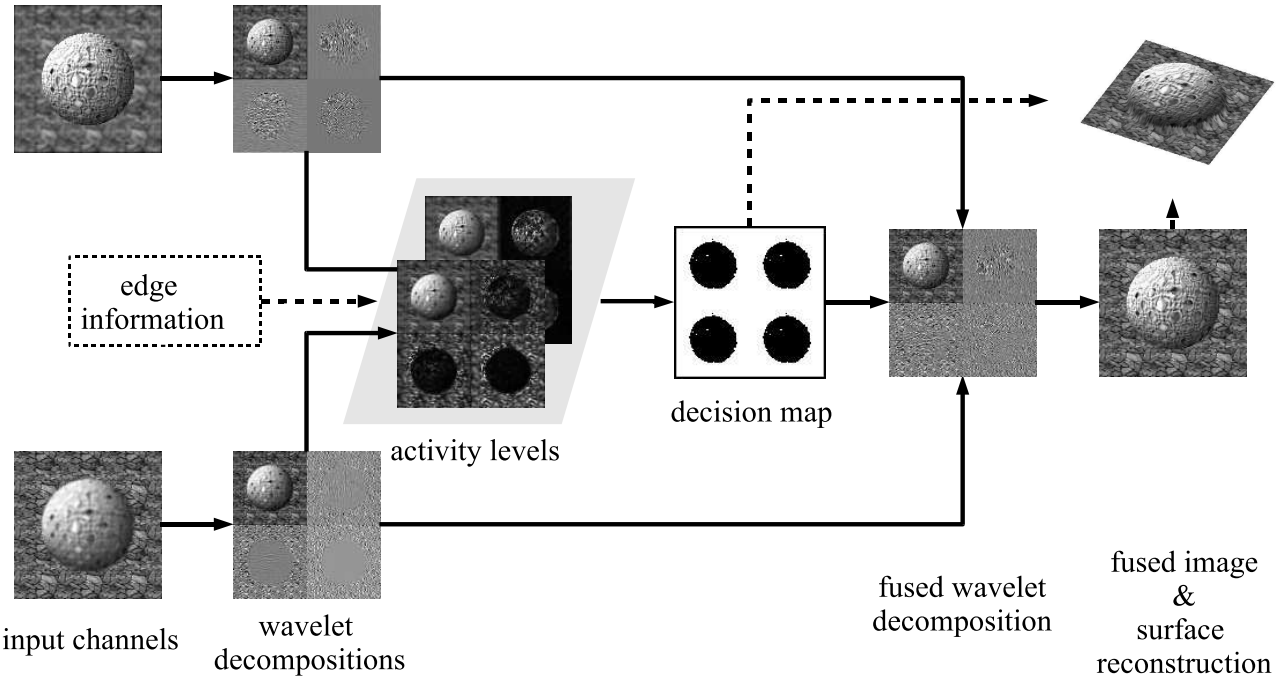


Figure 1. Multifocus fusion steps: Acquire input channel $\mathbf{I}_j$ with different focus settings; perform multiscale decomposition W_j (this diagram illustrates the one-level wavelet transform); calculate activity levels A_j; determine the decision map D using the maximum rule; combine the multiscale decompositions using the decision map and create a fused multiscale decomposition $W_{\mathbf{Z}}$; perform inverse multiscale transform to obtain a fused image $\mathbf{Z}$; reconstruct the object surface from the decision map and use the fused image as a texture.

Instead of averaging one can use rank filters. The popular choice here is to pick the maximum absolute value on the given neighborhood as our AL. Another option is to segment the input channels and calculate one AL for each image segment. In (2) the weight w then corresponds to the characteristic function of the segment. This approach is referred to region-based activity measurement.

To build DM, the most common scheme is to apply the maximum rule to AL's. Formally, we can write

$$D(\vec{\mathbf{p}}) = \arg\max_j \big(A_j(\vec{\mathbf{p}})\big)\,. \tag{3}$$

The decision map D has the same structure as W or A. It contains indices of the input channels and determines which MST coefficients to use at what place. In addition, if the focus setting for each channel j is known, D corresponds to the depth map and it can be used for the surface reconstruction. Using the decision map, the composite MST representation $W_{\mathbf{Z}}$ of the fused image $\mathbf{Z}$ is given by

$$W_{\mathbf{Z}}(\vec{\mathbf{p}}) = W_{D(\vec{\mathbf{p}})}(\vec{\mathbf{p}}) \tag{4}$$

The maximum rule considers at each position $\vec{\mathbf{p}}$ only the strongest MST coefficient and thus only one channel. Another possibility is to perform weighted averaging of the MST coefficients using weights proportional to AL's. However in the case of multifocus fusion, this combination scheme lacks any scientific support. Since we assume that each position (pixel) in the original image is acquired undistorted in at least one channel, only the maximum rule sounds perfectly plausible. The same holds true for any coefficient grouping method. One should avoid different decisions at different levels l and frequency bands k of MST. Therefore, we implement only one-level wavelet decomposition (one low-pass and three high-pass bands), calculate A as an average of AL's of three high-pass bands and use this A also for the low-pass band.

3. ORIENTED WINDOWS

In the introduction section we have outlined the reasoning for pursuing more accurate decision maps and that it can be achieved by implementing oriented windows based on edge information, i.e., calculating AL's with weighted averages that are space-variant.

The edge information consists of the direction and strength of an edge at each position of the channel. We extract the information from the low-pass band and adjust the weighting window w in (2) at each position as follows. Let $L(m,n)$ denote the low-pass band averaged over all the input channels, i.e., $L(m,n) = (1/J)\sum_{j=1}^{J} W_j(m,n,1,l_{\max})$. At each position $\vec{\mathbf{r}} = (m,n)$ we take a small neighborhood $N(\vec{\mathbf{r}})$ (typically 7x7 or 9x9) and estimate the distribution of the gradient $\nabla L = (L_m, L_n)$. To perform this, it is convenient to consider a Gaussian window as the weighting window. Then using the maximum likelihood method, one can estimate the 2x2 covariance matrix $\mathbf{C}$ of the Gaussian window as

$$\mathbf{C}(\vec{\mathbf{r}}) = \frac{1}{|N|} \sum_{\vec{\xi} \in N(\vec{\mathbf{r}})} \begin{pmatrix} (L_m(\vec{\xi}))^2 & L_m(\vec{\xi}) L_n(\vec{\xi}) \\ L_m(\vec{\xi}) L_n(\vec{\xi}) & (L_n(\vec{\xi}))^2 \end{pmatrix} \tag{5}$$

The covariance matrix decomposes into $\mathbf{C} = \mathbf{V}\left(\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right)\mathbf{V}^T$, where λ_1 is the variance in the direction of the main principal component of gradients and λ_2 is the variance in the direction orthogonal. The edge is perpendicular to the main principal component and since we want to orient the Gaussian window along the edge the covariance matrix becomes $\mathbf{C}_\perp = \mathbf{V}\left(\begin{smallmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{smallmatrix}\right)\mathbf{V}^T$. In addition, the effective window size should dependent on the magnitude of the gradient. If the image on the given neighborhood is smooth (small gradient) any change of the depth is highly improbable and the weighting window can (and should) grow to reduce the noise effect. Therefore, we propose to multiply the covariance matrix with a product $\lambda_1\lambda_2$, i.e., $\lambda_1\lambda_2\mathbf{C}_\perp = \mathbf{C}$ and we arrive at the former covariance matrix (5). The weighting window thus takes the form

$$w(\vec{\mathbf{r}}, s, t) = \frac{1}{Z} \exp\left\{ -\alpha \begin{pmatrix} s & t \end{pmatrix} \mathbf{C}(\vec{\mathbf{r}}) \begin{pmatrix} s \\ t \end{pmatrix} \right\}, \tag{6}$$

where Z is a partition function that guarantees $\sum_{s,t} w = 1$ and α is a parameter that depends on the noise level in the input channels and adjusts the overall effective size of the window. In practice, to speed up the calculation of AL in (2) we limit the size of w typically to 11x11. One can see that if there is no prevailing edge direction in the neighborhood, λ_1 are λ_2 are same in magnitude and w becomes isotropic.

4. EXPERIMENTS

We demonstrate the performance of the proposed fusion improvement on two sets of data. The first data set consists of images acquired with a standard digital camera in a laboratory environment. Apart from blurred images we are able to acquire an image which is sufficiently sharp everywhere to approximate a "ground truth" image and estimate an ideal decision map. We can then calculate the percentage of incorrect decisions (PID) to evaluate the quality of DM and the percentage mean squared error (PMSE) to evaluate the quality of the fused image. The evaluation measures are define as follows:

$$\mathrm{PID} = 100\frac{N_e}{N_t}, \tag{7}$$

where N_e is the number of erroneous decisions in the calculated DM and N_t is the total number of decisions, i.e., the size of the image, and

$$\mathrm{PMSE} = 100\frac{\|\mathbf{Z} - \mathbf{G}\|}{\|\mathbf{G}\|}, \tag{8}$$

where $\mathbf{G}$ is the "ground truth" image.

The second data set is more related to the application area and demonstrates the possible usage of the proposed method in microscopy.

We have tried various wavelets and studied the influence of the wavelet length. Short wavelets are too sensitive to noise while long wavelets do not provide enough discrimination power. Nearly symmetric orthogonal wavelets (8 coefficients in length) proposed in Ref.[9] seem to be a good compromise. We use them in all of the following experiments, in which we compare the proposed method of oriented windows with the standard approach of uniform squared windows. In the case of uniform windows the only adjustable parameter is the size of the window, which depends on the level of noise and size of details in the input images. In the case of oriented windows the similar role plays the parameter α that is adjustable. For the calculation of the edge information we set the neighborhood N to 9x9 in all the experiments.

The first experiment considers a simple two-plane scene with a Indian figure in front of a photography. Two images (Figs. 2(a)-(b)), one with the Indian in focus and the other one with the background in focus, were acquired with a standard digital camera and they act as input channels for fusion. The "ground truth" image in Fig. 2(c) was taken with a much larger aperture so it is in focus everywhere. In order to obtain an "ideal" decision map in Fig. 2(d), we replace the background photography with a white paper and take a picture that was easy to segment. Results of image fusion using oriented and uniform windows together with discrepancies in the decision maps are shown in Fig. 3. It is evident that the decision map for oriented windows is more accurate (PID = 4.8% in comparison with 6.2% in the case of uniform windows) however the visual improvement of the fused image is negligible.

To evaluate the noise robustness of the fusion technique, we added white Gaussian noise of different strength to the input images and calculated PID and PMSE. Results are summarized in Fig. 4. Clearly, the performance decreases as the noise level increases (signal to noise ratio - SNR decreases) but the method of oriented windows shows a constant performance boost in comparison with uniform windows.

In the second experiment, the proposed image fusion method was tested on multifocus images of a unicellular water organism called radiolarium. The radiolarium stack contains 51 images acquired by an optical microscope with the focal step of 1μm. An example of three images from the multifocus stack is in Fig. 5(a). The calculated decision map is in Fig. 5(c) and the corresponding fused image is in Fig. 5(b). Fig. 6 shows the 2.5D reconstruction of radiolarium with a surface derived from the decision map and the fused image as a texture. Since the reconstructed surface is a piece-wise constant approximation, higher-order surface interpolation was necessary to obtain the above result. Differences in the surface reconstruction between the proposed and standard method are apparent for close-ups in Fig. 7. The surface reconstructed with oriented windows matches the structure of radiolarium more accurately.

5. CONCLUSIONS

We have proposed to apply space-variant and edge-oriented window neighborhoods for the computation of activity levels in wavelet-based fusion in order to obtain more accurate decision maps. Presented experiments justify the proposed enhancement. The advantage of more accurate decision maps is fully exploited in the 2.5D surface reconstruction. Here, any outliers can create unrealistic peaks and valleys on the reconstructed surface and therefore accurate decision maps are crucial for a successful reconstruction.

ACKNOWLEDGMENTS

This work has been partially supported by the following grants: TEC2004-00834; the IM3 medical imaging thematic network from the Instituto de Salud Carlos III, the bilateral project: 2004CZ0009 CSIC-Academy of Sciences of the Czech Republic, and No. 102/04/0155, No. 202/05/0242 of the Grant Agency of the Czech Republic. F. Šroubek was also supported by the NATO Science fellowship.

REFERENCES

1. H. Wang, "A new multiwavelet-based approach to image fusion," *Journal of Math. Imaging and Vision* **21**, pp. 177–192, 2004.
2. M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Optical Eng.* **32**, pp. 2824–2836, 1993.
3. M. Subbarao and J. K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**, pp. 864–870, 1998.

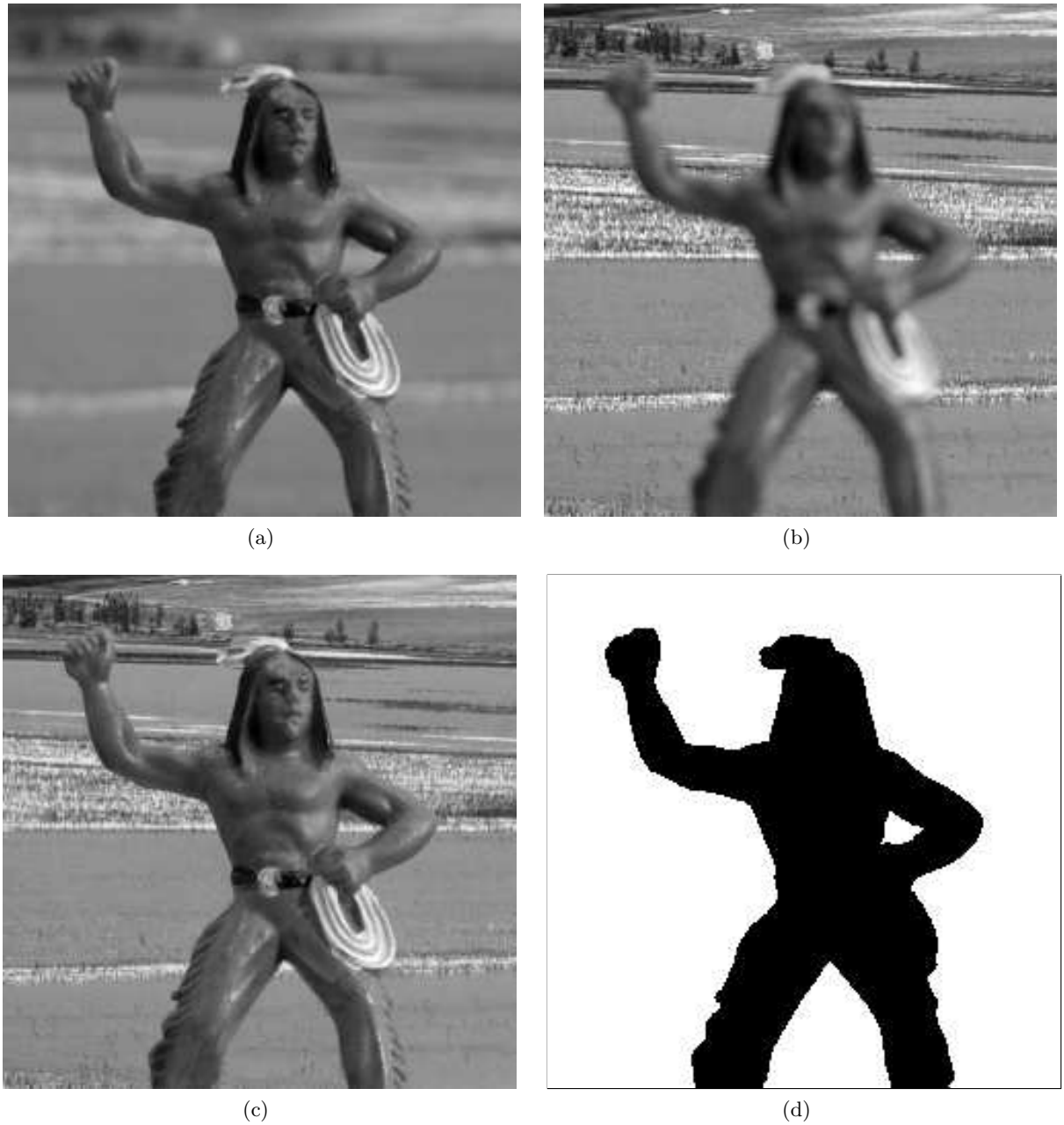


Figure 2. Laboratory experiment: Images depict a simple two-plane scene with a Indian figure in front of a photography. (a)-(b) Two input images with the Indian in focus and the background in focus, respectively. (c) Ideal image that is sharp everywhere was acquired with a large aperture. (d) Mask of the Indian that defines the ideal decision map.

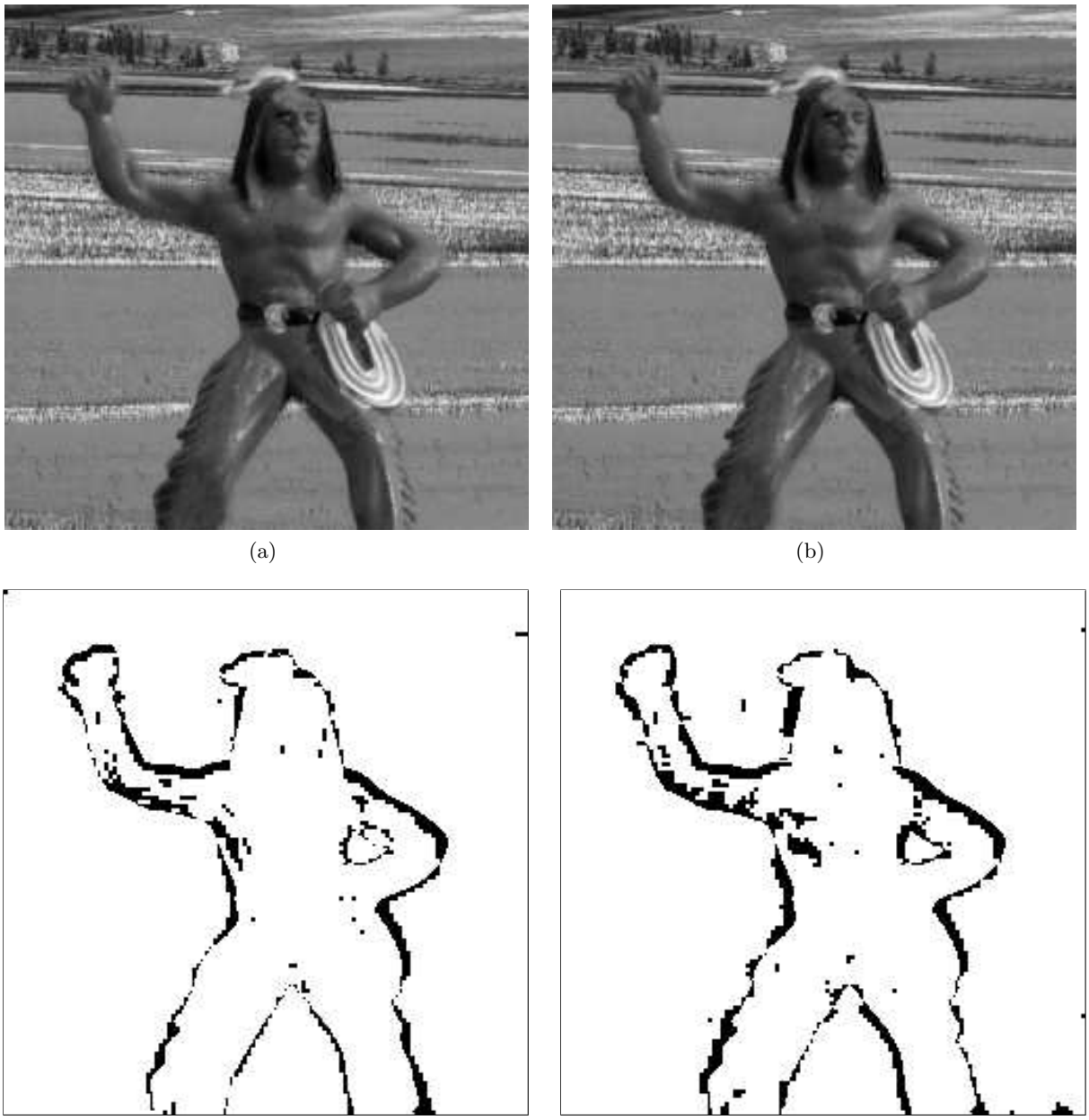
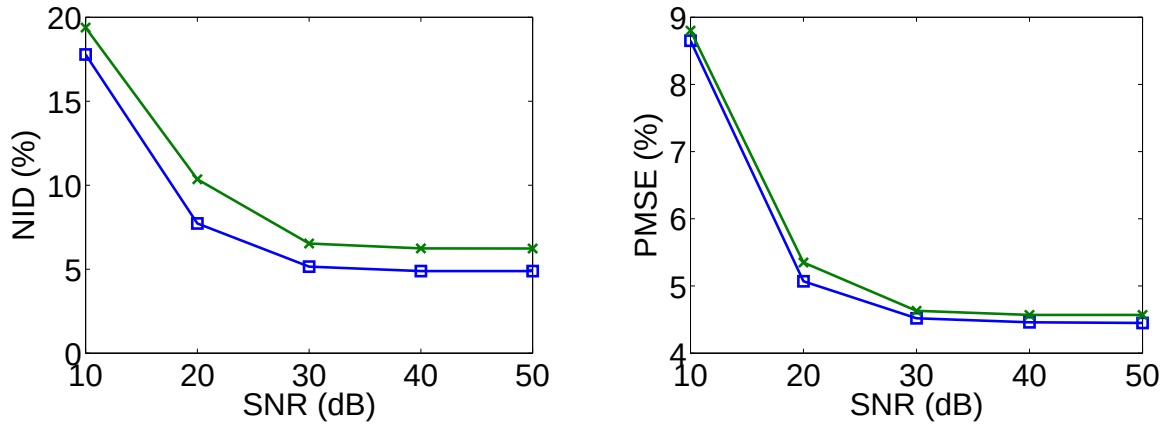
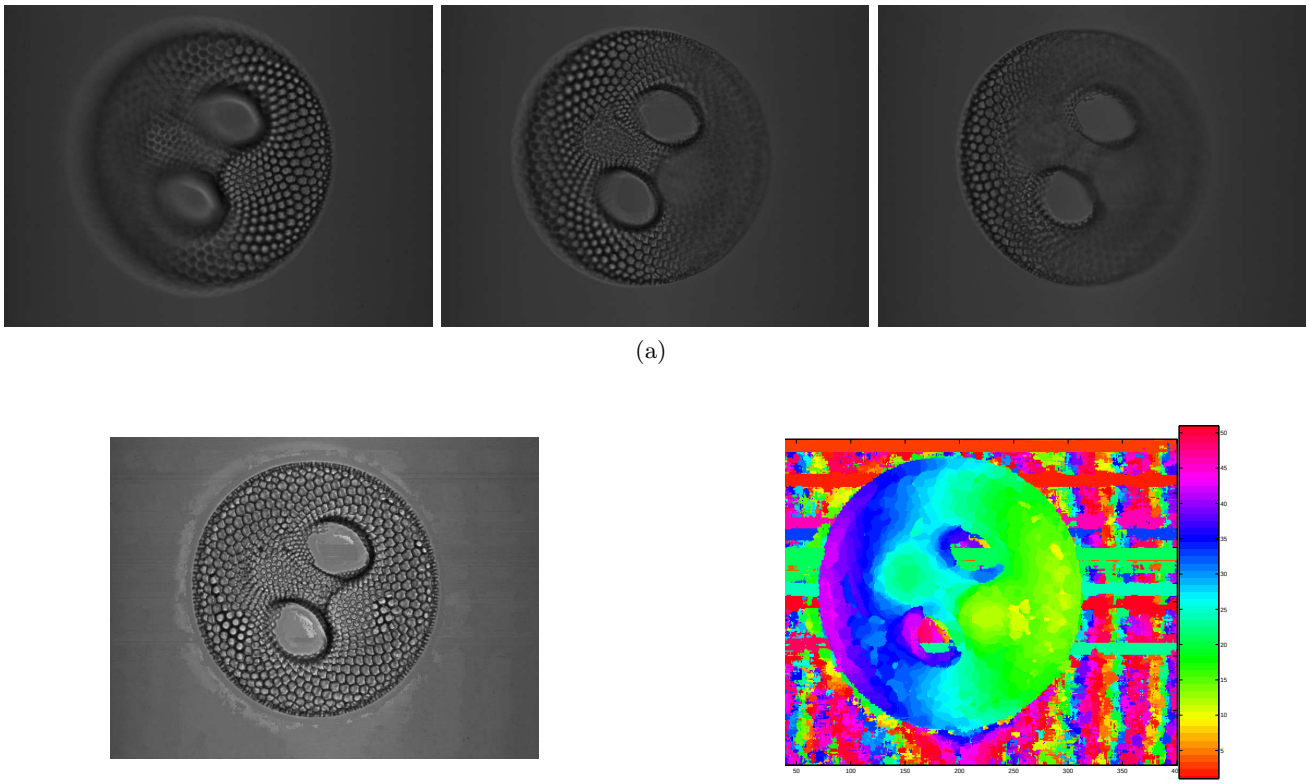


Figure 3. Laboratory experiment: Results of the wavelet-based fusion using oriented windows (a) and using standard uniform windows (b). Discrepancies (in black) between the ideal decision map in Fig. 2(d) and the calculated decision maps are in (c) and (d), respectively.



(a) (b)

Figure 4. Laboratory experiment: Performance of the wavelet-based fusion under different noise levels; (a) percentage of incorrect decisions (PID), (b) percentage mean squared error (PMSE). Comparison between uniform windows (×) and oriented windows (□).



(a)

(b) (c)

Figure 5. Biological application:(a) Example of three input images out of 51 images in the multifocus stack acquired with a optical microscope. (b) Fused image using the proposed wavelet-based approach with oriented windows. (c) Decision map calculated during the fusion procedure (image labels are colored according the color bar on the right).

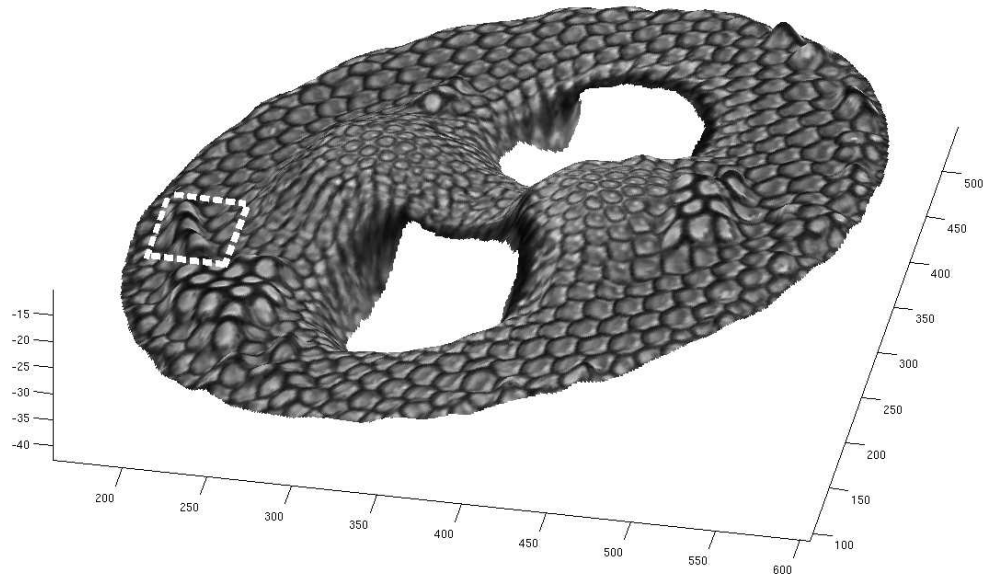
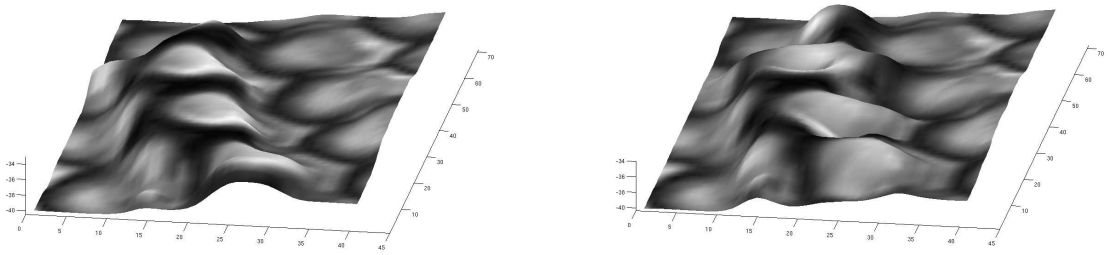


Figure 6. Biological application: Reconstruction of the object surface using the decision map in Fig. 5(c) as a depth map and the fused image in Fig. 5(b) as a texture.



(a) (b)

Figure 7. Biological application: Comparison of the surface close-up in Fig. 6 corresponding to the boxed region, calculated with oriented windows (a) and with uniform windows (b). Results in (a) clearly resembles the perceptual appearance of the surface better than in (b).

4. Y. Zhang, Y. Zhang, and C. Wen, "A new focus measure method using moments," *Image and Vision Computing* **18**, pp. 959–965, 2000.
5. H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Model and Image Processing* **57**, pp. 235–245, May 1995.
6. Z. Zhang and R. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," in *Proceedings of the IEEE*, **87**, pp. 1315–1326, Aug. 1999.
7. J. Kautsky, J. Flusser, B. Zitová, and S. Šimberová, "A new wavelet-based measure of image focus," *Pattern Recognition Letters* **23**, pp. 1785–1794, 2002.
8. G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," *Information Fusion* **4**, pp. 259–280, 2003.
9. A. Abdelnour and I. Selesnick, "Nearly symmetric orthogonal wavelet bases," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, May 2001.