Information Fusion xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Multifocus image fusion using the log-Gabor transform and a Multisize Windows technique

R. Redondo [a,*], F. Šroubek [b], S. Fischer [a], G. Cristóbal [a]

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

ARTICLE INFO

Article history:
Received 13 July 2006
Received in revised form 3 July 2007
Accepted 7 August 2008
Available online xxxx

Keywords:
Multiresolution fusion
Wavelet transform
Gabor representation
Overcompleteness
Additive noise
Multifocus imaging
Microscopy

ABSTRACT

Today, multiresolution (MR) transforms are a widespread tool for image fusion. They decorrelate the image into several scaled and oriented sub-bands, which are usually averaged over a certain neighborhood (window) to obtain a measure of saliency. First, this paper aims to evaluate log-Gabor filters, which have been successfully applied to other image processing tasks, as an appealing candidate for MR image fusion as compared to other wavelet families. Consequently, this paper also sheds further light on appropriate values for MR settings such as the number of orientations, number of scales, overcompleteness and noise robustness. Additionally, we revise the novel Multisize Windows (MW) technique as a general approach for MR frameworks that exploits advantages of different window sizes. For all of these purposes, the proposed techniques are firstly assessed on simulated noisy experiments of multifocus fusion and then on a real microscopy scenario.

1. Introduction

Three fusion levels of abstraction are usually identified. The lowest possible abstraction level is the pixel-level, which makes use of such physical parameters as the intensity value of the pixels. One step further, feature-level fusion, operates on attributes such as size, shape, edge, contrast or texture. Decision-level fusion is the last level of abstraction and deals with symbolic representations. A wide variety of mathematic tools that perform image fusion have been proposed in the literature [1]. This includes gray-value variance, averaging-PCA, neural networks, Bayesian modeling, non-linear filtering, Markov modeling and last, but not least, multiscale or multiresolution (MR) transforms.

Nowadays, MR transforms can be considered as the most popular tool for image fusion. MR-based image fusion lies between the pixel and feature-level, whose coefficients, often regarded as simple features, are represented in a joint space–frequency domain. Well known examples of MR transforms used in image fusion are the Laplacian pyramid [2], contrast pyramid [3], gradient pyramid [4], morphological pyramid [5], ratio-of-low-pass pyramid [6] and wavelet decomposition [7,8]. The *saliency measure*, also referred to as activity in the literature, constitutes a critical point in the whole process. Different methods were suggested, but in most cases the activity is proportional to the averaged energy of the frequency coefficients, where salient features (features to preserve) have high energy in high frequencies [9,10]. Thus, the high to mid-high frequency terms contain important details for our visual perception and understanding of the fused image. In multifocus fusion, for instance, the out-of-focus regions obviously lack high-frequency features. The norm of image gradient [4], norm of image Laplacian [11], energy of the Fourier spectrum [12], image moments [13] and energy of high-pass bands of wavelet transforms [8,14] are among the most popular activity measures.

Subsequently, image fusion usually relies on partitioning of the MR domain into subregions in which the saliency of features is measured. Averaging square neighborhoods (windows) is the simplest and the most common strategy for this. Then the activity of the regions is compared among all channels, and MR coefficients with the highest activity are preserved (*maximum selection rule*). By this maximum selection criterion an index matrix called a *decision map* is built, which identifies which coefficient to keep at each location. By tracking such indexes a single composite MR representation is built, whose inverse corresponds to the fused image. Instead of using maximum selection, more general approaches match coefficients by means of a certain *similarity measure*. Piella [15] offers an excellent overview of MR-based image fusion techniques.

Although neither evidences of superiority among MR decompositions nor among MR settings have been found, the literature, however, lacks an exhaustive comparison, mostly due to the fact that fusion evaluation cannot be always expressed in terms of

* Corresponding author.
E-mail address: rafa@optica.csic.es (R. Redondo).

1566-2535/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.inffus.2008.08.006

objective criteria. Consequently, multiple perceptual and statistical evaluation metrics have been suggested [16–18].

The novel log-Gabor implementation that we recently proposed elsewhere [19] opens doors to those image processing applications which require exact reconstruction. Hence, it has already shown notable performance in image restoration and image compression [20,21]. One objective of the current paper, therefore, involves evaluating the log-Gabor wavelets as a novel MR transform for the field of image fusion. In addition, it is a well known fact that the appropriate averaging window size is unfortunately highly dependent on the scale of objects and the noise present in the source images. Whereas large windows improve robustness, small ones perform accurate space localization. The *Multisize Windows* (MW) technique that we sketched elsewhere [22] locally adapts the size of the window according to the local features in the image, exploiting advantages of manifold sizes. Although it was evaluated under a wavelet framework as a first attempt, its formulation is valid for any MR scheme.

In light of the background information mentioned above, we established the following objectives for this work: (1) to carry out a quantitative evaluation of the log-Gabor wavelets; (2) to configure MR settings for successful image fusion; (3) to create a novel MW technique with the log-Gabor decomposition; and (4) to compare different wavelet-based fusions. To deal with these challenging goals, the course of the paper is as follows: MR-based settings are described in Section 2, log-Gabor wavelets and some MR features are discussed in Section 3, the MW technique is reviewed in Section 4 and finally experiments for two multifocus scenarios are presented in Section 5.

2. A MR-based fusion scheme

2.1. MR formulation

We follow in part the notation and terminology given in [15]. Let x_S denote the Sth input channel and y_S the MR transform of x_S, where the *analysis operator* Ψ satisfies $y_S = \Psi(x_S)$ and the *synthesis operator* Ψ^{-1} satisfies $x_S = \Psi^{-1}(y_S)$, assuming exact reconstruction. Let the index $\mathbf{n} = (n, m)$ address the MR coefficients in a sub-band with a given orientation $p = \{1, \ldots, P\}$ (vertical, horizontal and several diagonal filtering directions) and scale $k = \{1, \ldots, K\}$. The highest scale (lowest resolution level) $y_S(\mathbf{n}, 1, K)$ is made of a unique band usually called the residual or *approximation* and the remaining high-pass bands $y_S(\mathbf{n}, p, k)$ are referred to as the *detail pyramid*.

2.2. Activity measurement

MR coefficients reflect the energy of frequency terms in a given local area. The activity is thus related to the absolute or squared value of the coefficients. Because *sample-based* operation is rather vulnerable to the presence of noise in the channels, most approaches select a fixed window, performing an *area-based* operation, to average the activity as follows:

$$a_S^N(\mathbf{n}, p, k) = w^N(\mathbf{n}) * |y_S(\mathbf{n}, p, k)|, \tag{1}$$

where a_S^N is the activity, w^N is any type of normalized $N \times N$-window and the operator $*$ means the convolution operator. We assume windows of Gaussian type with a standard deviation equal to the window size, i.e. $\sigma^2 = N^2$.

2.3. Decision maps

The maximum selection rule, $a^N = \max_S[a_S^N]$, selects only the strongest MR coefficient among the channels at each location. The maximum sounds adequate in multifocus fusion because we assume that each pixel is acquired in-focus in at least one channel. Weighting or thresholding MR coefficients appears, in principle, more suitable for multisensor fusion. However, even then the maximum selection rule is often favored [8].

Despite the fact that space–frequency representations often accomplish efficient image decorrelation, most MR coefficients are highly correlated with other surrounding coefficients and with equivalent coefficients in adjacent sub-bands (scale and orientation) [23,24]. In order to make decision maps consistent in different bands as well, we extend the maximum selection rule across orientations:

$$A_S^N(\mathbf{n}, k) = \max_p \left[a_S^N(\mathbf{n}, p, k)\right]. \tag{2}$$

This tactic could also improve noise robustness whether or not a preferred oriented pattern occurs. Observe that it may also be similarly extended across scales (not implemented here).

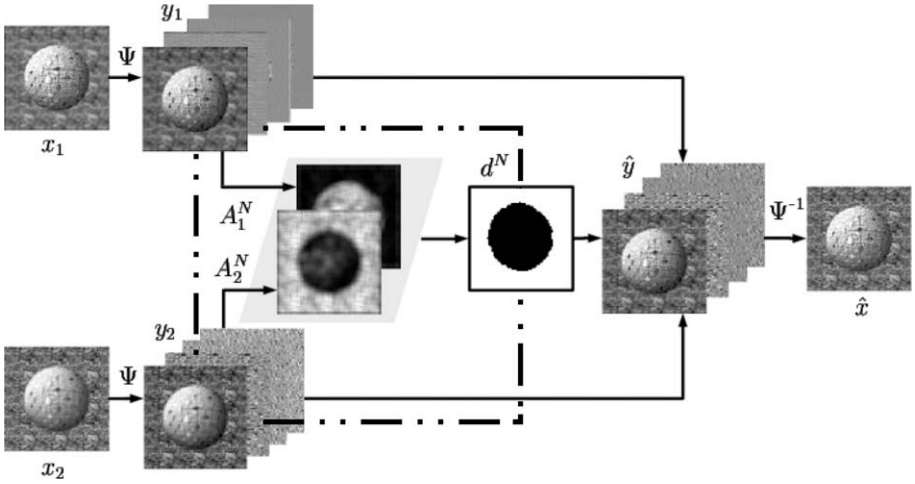

Fig. 1. MR-based (multifocus) fusion scheme: input channels x_1 and x_2 are acquired with different focus settings, MR decompositions y_1 and y_2 are performed with $K = 2$ scales and $P = 3$ orientations, and activity A_1^N and A_2^N are calculated by means of a given window N and by taking the maximum activity among orientations. Decision map d^N is obtained by using the maximum selection rule and the composite MR domain $\hat{y}$ is merged from the decided MR coefficients. Finally, the inverse MR transform yields the fused image $\hat{x}$.

Please cite this article in press as: R. Redondo et al., Multifocus image fusion using the log-Gabor transform and a Multisize Windows ..., Informat. Fusion (2008), doi:10.1016/j.inffus.2008.08.006

The decision map d is finally built by taking the argument of the maximum activity among channels:

$$d^N(\mathbf{n},k) = \arg\max_S \left[A_S^N(\mathbf{n},k)\right]. \quad (3)$$

Concerning the approximation $y_S(\mathbf{n},1,K)$, a typical procedure involves weighted averaging the low-pass residue of all input channels. That is specially important when images have been captured at different lighting conditions or with different sensors [15]. In multifocus fusion, however, all channels have a similar low-pass response and therefore can be simply averaged.

2.4. Fused image

Finally, the composite MR decomposition denoted as $\hat{y}$ is merged point-by-point as follows:

$$\hat{y}(\mathbf{n},p,k) = y_{d^N(\mathbf{n},k)}(\mathbf{n},p,k), \quad (4)$$

where the fused image $\hat{x}$ is obtained by the synthesis operator as $\hat{x} = \Psi^{-1}(\hat{y})$. The complete procedure is illustrated in Fig. 1.

3. The log-Gabor transform on image fusion

3.1. Log-Gabor formulation

Log-Gabor filters, illustrated in Fig. 2, consist of a complex-filtering arrangement in p orientations and k scales, whose expression in the log-polar Fourier domain is as follows:

$$G(\rho,\theta,p,k) = \exp\left(-\frac{1}{2}\left(\frac{\rho-\rho_k}{\sigma_\rho}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{\theta-\theta_{k,p}}{\sigma_\theta}\right)^2\right) \quad (5)$$

in which (ρ,θ) are the log-polar coordinates and $(\sigma_\rho,\sigma_\theta)$ are the angular and radial bandwidths (common for all the filters). The pair $(\rho_k,\theta_{k,p})$ corresponds to the frequency center of the filters, where the variables p and k represent the orientation and scale selection, respectively. In addition, the scheme is completed by a Gaussian low-pass filter $G(\rho,\theta,1,K)$ (approximation). In order to attain an exact reconstruction, further filtering parameters were tuned by following the design constraints recommended in [19].

3.2. Overcompleteness, complex nature and other properties

Because of their Gaussian profile, Gabor functions provide an optimum joint space–frequency localization whose shape is smooth, symmetric, infinitely differentiable and monomodal, without side lobes either in space or frequency domain [25]. Like most wavelet decompositions, the log-Gabor transform is built up as a dyadic structure in frequency octaves; moreover, it tolerates an arbitrary number of orientations (the second exponential term in Eq. (5)). Furthermore, its complex nature has shown advantages in image processing [26,27] since a coupled pair of filters with opposite phases (real/imaginary) is able to respond simultaneously to features of different phase, such as edges or ridges. Additionally, recent articles have paid attention to the disadvantages of critical wavelet sampling [10,15,28–30]. Although the typical dyadic filter bank of wavelets cancels deep overlapping (aliasing) that exists between sub-bands, a slight manipulation of the coefficients (shifting, rotation and scaling) can induce unpredictable artifacts. In some applications this may not be relevant, but in image fusion this is definitely undesirable as it may lead to misregistered images. One must be aware that overcompleteness unfortunately increases the dimensionality of the transform domain and therefore the computational cost and memory consumption.

4. The Multisize Windows technique

The MW technique was designed to mitigate the trade-off between robustness and localization. Large windows contribute to a more robust activity calculation but may result in feature cancellation or contrast loss, causing smoothness in the decision maps. In contrast, small windows improve localization but gravely impair the robustness. Previously introduced in [22], we contribute here a refined version of the technique.

4.1. MW formulation

Given a certain set $\mathcal{N}$ of averaging window sizes, the idea is to trust in the largest window wherever it does not impair localization. To measure such a localization impairment, a window confidence is defined proportionally to the highest number of neighboring decisions d^N that come from the same channel. We formally define it as

$$C^N(\mathbf{n},k) = \max_S \left[\frac{1}{N^2}\left|\{\mathbf{n}' \in \widetilde{N} : d^N(\mathbf{n}',k) = S\}\right|\right], \quad (6)$$

where the operator $|\cdot|$ means the cardinality of decisions indexing the same Sth channel and $\widetilde{N}$ means a neighborhood around n of size $N \times N$. Note that C^N runs from 0 (uncertain) to 1 (certain) and that this value can be also interpreted as the probability in a given neighborhood of size N that the coefficients $\hat{y}(\mathbf{n},\cdot,k)$ come from the channel S. Also note that for convenience we match up this neighborhood with the size N of the averaging window referred to in Eq. (1). A large window is then applied with high confidence in cases when most decisions (coefficients) point to (should be taken from) the same channel. On the other hand a smaller window should be used when disparity in the decisions is manifested. The entire procedure can be described as follows:

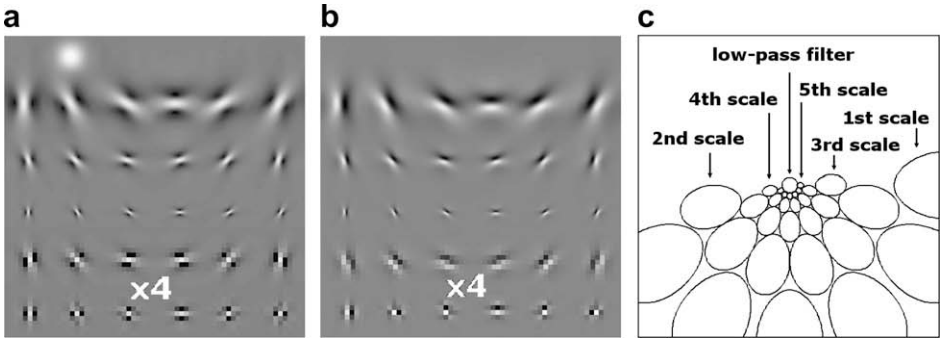

Fig. 2. Multiresolution log-Gabor arrangement with $P = 6$ orientations and $K = 6$ scales. (a) Real part. The low-pass filter is drawn in the upper-left part. (b) Imaginary part. (c) Filters in the Fourier domain.

Please cite this article in press as: R. Redondo et al., Multifocus image fusion using the log-Gabor transform and a Multisize Windows ..., Informat. Fusion (2008), doi:10.1016/j.inffus.2008.08.006

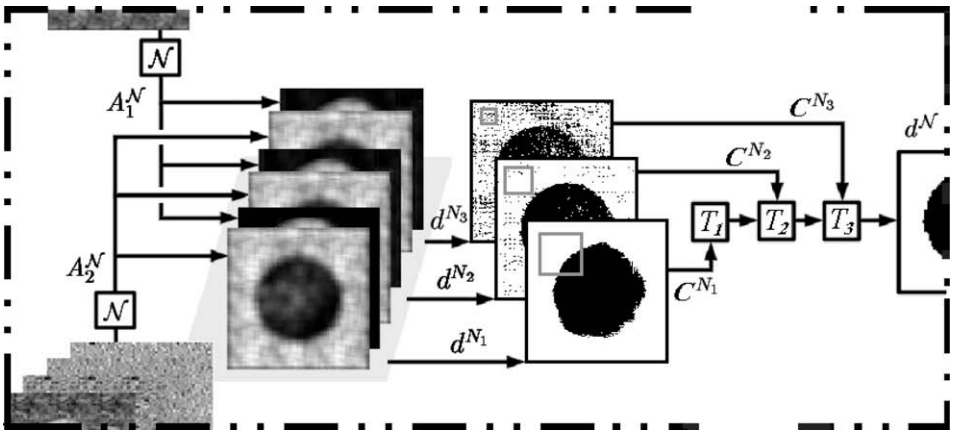

Fig. 3. Multisize windows scheme (one scale is shown): activity maps $A_1^{\mathcal{N}}$ and $A_2^{\mathcal{N}}$ are obtained by applying the set of windows $\mathcal{N} = \{N_1, N_2, N_3\}$; three decision maps are then released. The MW technique is hierarchically computed from the largest to the smallest window according to three given thresholds in $\mathcal{T} = \{T_1, T_2, T_3\}$. The obtained decision map $d^{\mathcal{N}}$ merges, as usual, the composite MR domain. The same procedure would be similarly applied in subsequent scales. Note that the dashed square inside Fig. 1 corresponds to the dashed frame surrounding the current figure.

(1) Let $\mathcal{N} = [N_1, N_2, \ldots, N_M]$ be a set of windows sorted by decreasing size, where $i = 1, \ldots, M$ indexes the window sizes;
(2) Let $\mathcal{T} = [T_1, T_2, \ldots, T_M]$ be a threshold set for each $N_i \in \mathcal{N}$, where $T_i \in [0, 1]$;
(3) Initialize $d^{\mathcal{N}}(\mathbf{n}, k) = 0$ as the MW decision map;
(4) For $i = 1$ to M
(5) For all $d^{\mathcal{N}}(\mathbf{n}, k) = 0$
(6) If $C^{N_i}(\mathbf{n}, k) \geqslant T_i$ then
(7) $d^{\mathcal{N}}(\mathbf{n}, k) = \arg[C^{N_i}(\mathbf{n}, k)]$;
(8) end

4.2. Some MW insights

Note that the above steps have to run from the largest to the smallest window. Note also that locations already assigned in $d^{\mathcal{N}}(\mathbf{n}, k)$ at any previous step are not further computed. In the case that two different decisions are equally frequent, one of them is simply taken. This hierarchical combination, however, might be unable to merge all of the coefficients, i.e. some coefficients might be below T_i for any window N_i. In the previous proposal [22] a linear combination of activities derived from the whole windows set was allocated to those 'unmerged' coefficients. Alternatively, we include in $\mathcal{N}$ the smallest possible window $N_M = 1$ to satisfy the condition in step (6) for any threshold. Finally, the block diagram in Fig. 3 sketches the procedure and Fig. 4 presents a simple example to gain insight into the MW technique.

5. Performance evaluation in multifocus frameworks

Multifocus frameworks, tackled particularly in this paper, consist of a given stack of images capturing the same scene but with different focus planes. Thus objects are focused differently in the input images. The activity is often referred to as *focus measure* and the decision maps identify regions in focus. The ideal multifocus fused image is at best focused everywhere.

Hereafter the number of levels will be $K = 4$ (note that the fourth level is the approximation pyramid) and the number of orientations P will vary from 1 to 18 (from isotropic to highly anisotropic). A broad window set will be chosen $\mathcal{N} = [19, 15, 11, 9, 7, 5, 3, 1]$ whose set of confidence thresholds will be $\mathcal{T} = [0.8, 0.8, 0.8, 0.7, 0.7, 0.6, 0.6, 0.6]$. The thresholds of small windows were relaxed heuristically because such small windows perform last in the MW procedure and precisely in regions of transitions between objects located at different focus planes where these small windows are the most reliable in size.

5.1. The laboratory scenario

The data set consists of two images acquired with a conventional digital camera in a laboratory environment. Multiexposure and misregistration are practically cancelled. Apart from the blurred versions in Fig. 5a and b (tower and background in focus, respectively), we constructed a very precise reference map by putt-

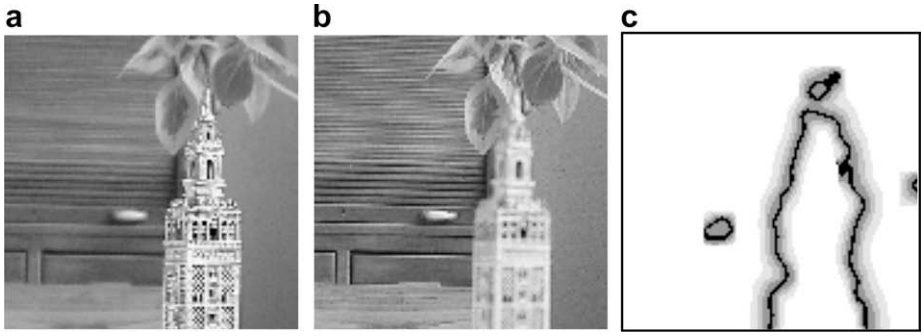

Fig. 4. A graphical example showing the adaptation of the window size performed by the Multisize Windows technique. (a, b) Two input images of the same scene with different focus planes, the tower and the background are in focus, respectively. (c) White pixels indicate those locations where the largest window was applied. Then, the darker the pixel, the smaller the window that was applied (five windows were chosen). Note that the importance of the windows (confidence) fluctuates close to the focus plane transition.

Please cite this article in press as: R. Redondo et al., Multifocus image fusion using the log-Gabor transform and a Multisize Windows ..., Informat. Fusion (2008), doi:10.1016/j.inffus.2008.08.006

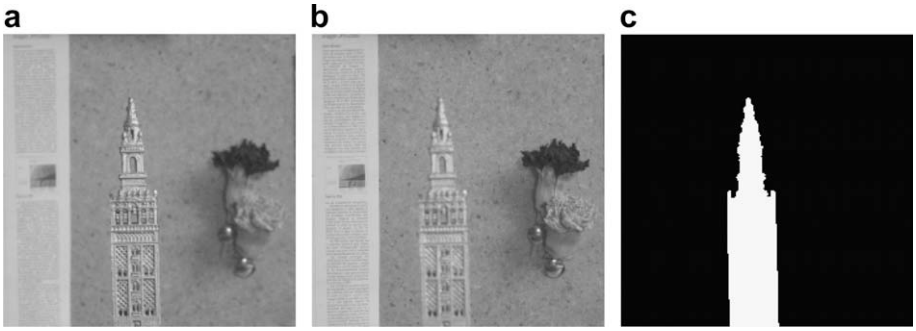

Fig. 5. Laboratory experiment. (a, b) A two-plane scenario made of two images of size 236×236 with the tower and the background in focus, respectively. (c) Ideal decision map constructed by occluding the background with a piece of black felt and subsequent gray-value thresholding.

ing a piece of black felt between both focus planes and applying a simple threshold algorithm, as shown in Fig. 5c. In order to evaluate the quality of the decision maps, the *percentage of errors* (E) is defined as follows:

$$E = 100\frac{n_e}{n_t}, \tag{7}$$

where n_e is the number of wrong decisions and n_t is the total number of decisions, i.e. the size of the image.

The two focus planes in Fig. 5 are deliberately placed not too far away from each other in order to impair focus estimation and therefore to better discriminate among algorithms. We believe the proposed scenario is complex enough, in the sense of object shapes, textures and blurring, to elucidate several questions regarding appropriate MR settings. Two focus planes facilitates the exercise of constructing an ideal reference and therefore a quantitative measure. Moreover, it permits straightforward control of what is happening, whereas a large number of planes would complicate the process.

5.1.1. Size of the window

The appropriate window size is conditioned by the size/shape of the objects and the noise present in the channel. In Fig. 6, three examples are shown as representative samples of the influence of the window size in decision maps. Fig. 6a illustrates that small windows are more exposed to noise and prone to errors in the frequency analysis. Hereafter we will call them *spurious errors*. On the other hand the accurate outline of the tower verifies the capability of spatial localization. Conversely in Fig. 6b, a bigger window yields robust analysis where most errors are committed close to focus plane transitions. Hereafter we will call them *localization errors*. The result obtained through the MW technique in Fig. 6c presents a notable decrease in both types of errors. Large windows were applied mostly on inner regions of the focus planes, removing spurious errors and smaller windows improved localization around focus plane transition, such as the tower's contour.

Several statements can be made from the graph in Fig. 7. Windows of size $N < 5$ commit a large amount of spurious errors. For $N > 11$ coefficients increase the amount of localization errors until the errors are clearly significant/visible. Therefore, it seems that the best window size ranges from $N = 5$ to $N = 11$. It can be said that in this 'optimum' range the amount of spurious errors and the amount of localization errors are well balanced. As the decomposition level is higher, that behavior shifts to bigger window sizes,

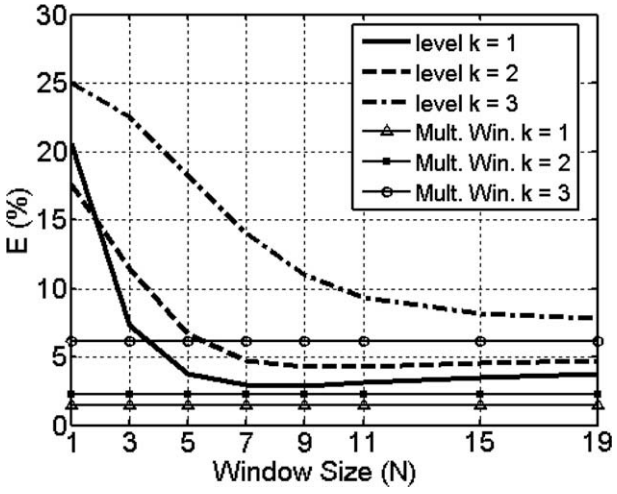

Fig. 7. Percentage of errors (E) committed in decision maps by the log-Gabor transform ($P = 3$) vs. window size (N) and for the Multisize Windows for the three levels of decomposition ($k = \{1, \ldots, 3\}$).

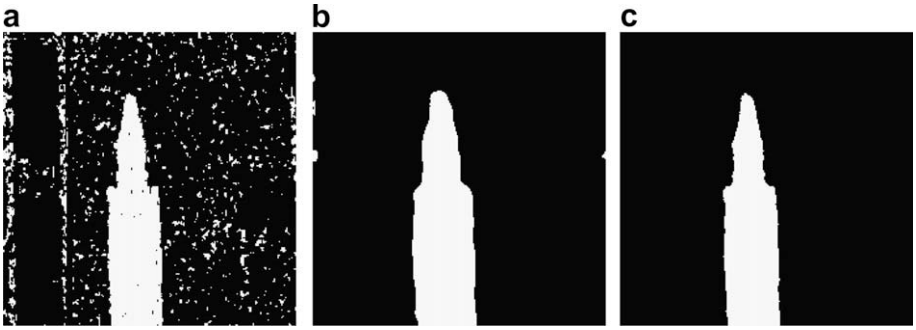

Fig. 6. Decision maps obtained through log-Gabor ($k = 1$ and $P = 3$) with a window of size (a) 3×3 ($E = 7.29\%$), (b) 15×15 ($E = 3.52\%$), and (c) the Multisize Windows technique ($E = 1.44\%$).

Please cite this article in press as: R. Redondo et al., Multifocus image fusion using the log-Gabor transform and a Multisize Windows ..., Informat. Fusion (2008), doi:10.1016/j.inffus.2008.08.006

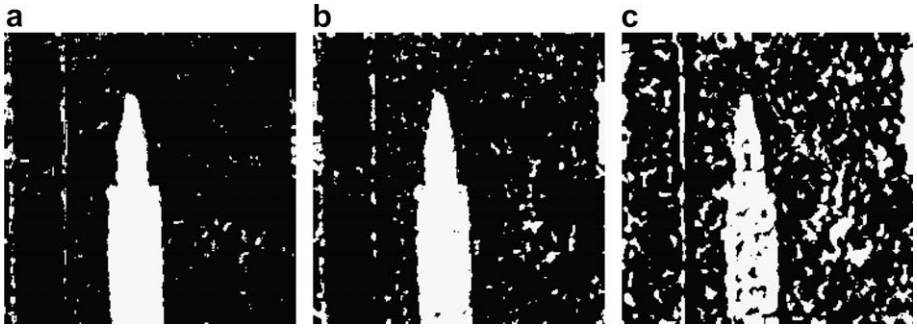

Fig. 8. Decision maps obtained through log-Gabor ($P = 3$ and $N = 5$) and level (a) $k = 1$ ($E = 3.8\%$), (b) $k = 2$ ($E = 6.71\%$) and (c) $k = 3$ ($E = 18.17\%$).

i.e. higher decomposition levels require bigger window sizes. The lowest E was delivered by the MW technique regardless of window size, the decomposition level or number of orientations (not shown here).

5.1.2. Number of decomposition levels

The degree of blurring present in the channels determines the required number of decomposition levels. The maps in Fig. 8, together with the plot in Fig. 7, confirm that levels above $k = 3$ are unnecessary and even counter-productive due to the high presence of spurious errors as well as poor localization. This is again confirmed across orientations and is independent of the window size. For simplicity, we will consider hereafter only the first level of decomposition, i.e. $k = 1$.

5.1.3. Number of orientations

Again several conclusions can be stated from the plot in Fig. 9. If the windows are small (i.e. $N < 5$), increasing the number of orientations contributes to reducing spurious errors. Unfortunately, from medium to big sized windows this action does not yield significant improvements. Thus, this plot justifies the use of sizes $N \geqslant 5$. Surprisingly, the no orientations condition ($P = 1$) delivers notable performance for $N > 3$. The MW technique shows its robustness across orientations. Note that, though the impairment is high for the smallest windows, most decisions come from the first stages in the MW procedure, that is, for the biggest windows, and therefore decision maps are not largely affected.[1]

5.1.4. Noise robustness

The previous data set in Fig. 5 was artificially corrupted with additive noise of Gaussian probability distribution [31] which characterizes many natural phenomena. Noise energy is characterized by its variance, which ranges for this and subsequent studies from imperceptible to very harsh noisy conditions (far from normal working conditions). Each setting was averaged among 30 noise instantiations and the variance encountered was so negligible as to be discarded.

The performance against noise is plotted in Fig. 10, from which several conclusions can be made. Medium window sizes perform better for low noise ($< 10^{-5}$ in this case), as in the absence of noise in Fig. 7. Here, differences in E are minor but still visible, as in Fig. 6. As the noise levels increase spurious errors diminish, up to a certain limit, according to the area of the window. The MW behaves with an elevated resistance to noise since it delivered the lowest E against the overall set of single-window approaches.

For medium and large windows ($N \geqslant 5$), no significant gain in noise robustness was encountered from the use of a high number of orientations (not shown). For small windows ($N < 5$) the use of more orientations makes E diminish only for low noise conditions ($< 10^{-5}$ in this case), however, decision maps are still largely corrupted.[2] These results agree with previous observations in the absence of noise (Fig. 9).

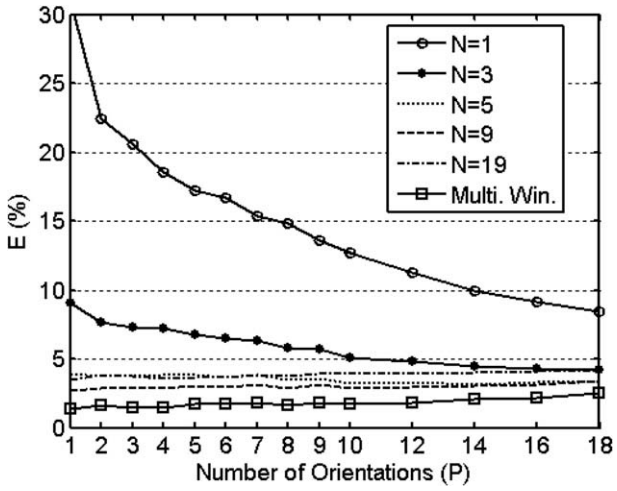

Fig. 9. Percentage of errors (E) committed in decision maps by the log-Gabor transform ($k = 1$) vs. the number of orientations (P) for different window sizes (N) and for the Multisize Windows technique.

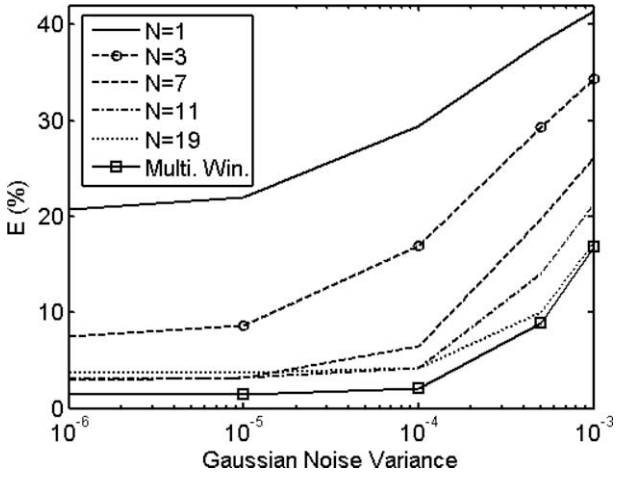

Fig. 10. Percentage of errors (E) committed in decision maps by the log-Gabor transform ($k = 1$ and $P = 3$) against increased variance of Gaussian noise for different window sizes (N) and the Multisize Windows technique.

[1] This observation was also corroborated for steerable filters.

[2] This observation was also corroborated for steerable filters.

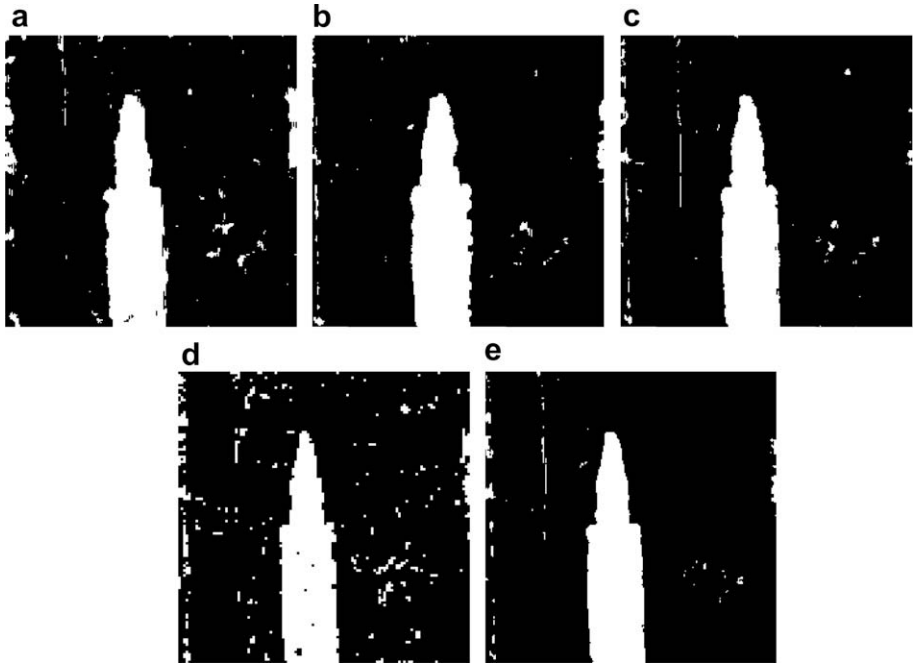

Fig. 11. Decision maps (in absence of noise) obtained through (a) steerable ($E = 4.93\%$), (b) Daubechies (db4) ($E = 3.94\%$), (c) biorthogonal (bior4.4) ($E = 3.26\%$), (d) decimated biorthogonal (bior4.4) ($E = 5.11\%$) and (e) log-Gabor wavelets ($E = 2.93\%$). All wavelets were equally configured with $k = 1$, $P = 3$ and $N = 7$, except for the decimated biorthogonal filter ($N = 3$).

5.1.5. Type of wavelet

Among all wavelet families existent in the literature, the following were selected as a representative sample according to their nature: *Daubechies*' filters (non-symmetric, orthogonal in principle but undecimated), *steerable* filters (symmetric and non-orthogonal), *biorthogonal* filters (symmetric and undecimated) and *log-Gabor* (complex-valued, symmetric and non-orthogonal). A decimated version of biorthogonal filters was also considered.[3] In order to avoid possible bias these filters were tuned to the same best setting encountered in previous sections: $k = 1, P = 3$ and $N = 7$ (the decimated version used $N = 3$).

From the five decision maps (in absence of noise) depicted in Fig. 11 it can be said that the decimated biorthogonal filters committed a lot of spurious errors and localization errors (the tower was poorly outlined); the steerable filters performed the worst among the undecimated versions, closely followed by the Daubechies filters and finally by the undecimated biorthogonal version, which yielded a similar quality that was a bit worse than the log-Gabor. Note that these decision maps still maintain a high E in comparison to the output of MW, which achieved, by far, the best result in terms of E as well as visual quality (see Fig. 6c). Such differences in performance were also observed for other window sizes and decomposition levels (not included here).

The rates plotted in Fig. 12 were also replicated for speckle and salt and pepper distributions [31] (not shown here). The results were found to be very similar to the Gaussian case. The following can thus be concluded: all wavelets seemed to be proportionally affected by noise but they respond equally against high noise level; decimation severely impairs robustness and the quality of the decision map; steerable filters committed about 2% more errors than other wavelets[4]; biorthogonal and log-Gabor wavelets performed almost equally well and were the best in the absence of noise and in low-noise conditions, the MW technique outperformed, by far, any single-window approach (although it was only plotted in combination with log-Gabor as an example of one of the best rated). A similar improvement degree was observed for all types of wavelets by using the MW. Moreover, the MW technique improved noise robustness above 5% at 10^{-4} and above 10% at 10^{-5}, probably because of the benefit of using large windows at the high noise level.

5.2. A microscopy scenario

Multifocus fusion is especially relevant in microscopy where specimens can not be acquired completely in focus with a single image due to the narrow depth-of-field. As with going beyond physical limitations, extending the depth-of-field has been a challenging pursuit since the pioneering research on microscopy in the early 1980s [27,33–36].

An applied fusion of images of a real sample of a fly taken from a bright-field microscope is depicted in Fig. 13a–d. This sample contains complicated structures as thin eyelashes which give rise to abrupt focal plane transitions. Despite the fact that most high-frequency details have been merged with $K = 3$, an overall blurring is visible in comparison to $K = 4$ (see Fig. 13e and f). It can be also observed in Fig. 13g–i that the number of orientations, in spite of not having a special impact on decision maps, helps to preserve the continuation of some oriented features such as the eyelashes. The MW result in Fig. 13j also offers a slight improvement on the oriented features. Furthermore, it was observed (not shown here) that an elevated number of orientations ($P > 10$) did not improve the fused images at all but rather some blurring artifacts could appear. Whereas decimated wavelets introduce very strong ringing, undecimated biorthogonal filters yielded one of the best visual qualities (see Fig. 13k and l). Low-frequency artifacts, however, still arose along the eyelashes, but they are attenuated by the log-Gabor approach, albeit accompanied by some ringing in Fig. 13m. The MW technique in Fig. 13n improves some detailed features but the overall visual quality remains similar to the best single-window log-Gabor approach.

[3] A fair error calculation entails taking a half-size window and decimating the ideal decision map in Fig. 5c.

[4] Although the high-pass residue was firstly discarded, a new 'steerable' residue proposed in [32] was also tested but did not deliver better results.

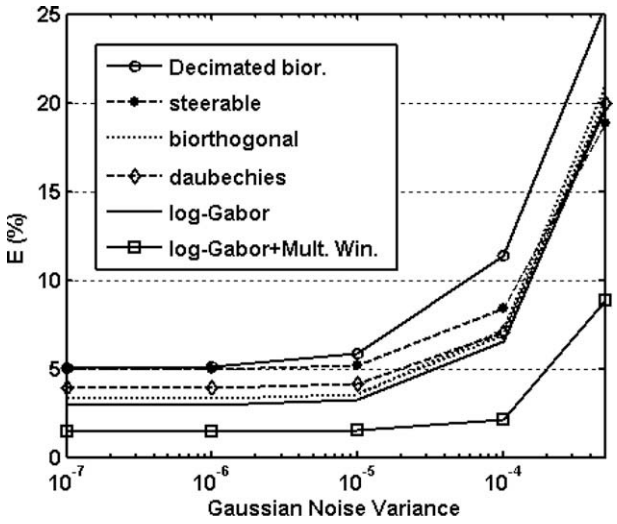

Fig. 12. Percentage of errors (E) committed in decision maps vs. Gaussian noise for different types of pyramid decompositions. All wavelets were equally configured with $k = 1, P = 3$ and $N = 7$, except for the decimated biorthogonal filter ($N = 3$).

5.3. Complexity

It is well known that, taking n as the total number of pixels, the complexity of the FFT used for computing log-Gabor filters is $O(n \log n)$ [19] and the memory slots required for the undecimated domains are $n \times K \times P$, whereas the complexity of the lifting schemes is $O(n)$ [37] and the number of memory slots is n. In addition, both the single-window and the MW approaches require a fixed number of operations for each pixel; hence their complexity is the same, i.e. $O(n)$. But the number of operations also increases linearly with the window area N^2 for the single-window approach and with the sum of all windows areas for the MW method. It is then advisable to limit the number of large windows in the MW approach. Interestingly, MW does not correspondingly increase its memory requirement since the same memory allocation can be used in the hierarchical computation.

6. Conclusions

We aimed to evaluate log-Gabor filters and a whole range of parameters involved in image fusion not only by means of quanti-

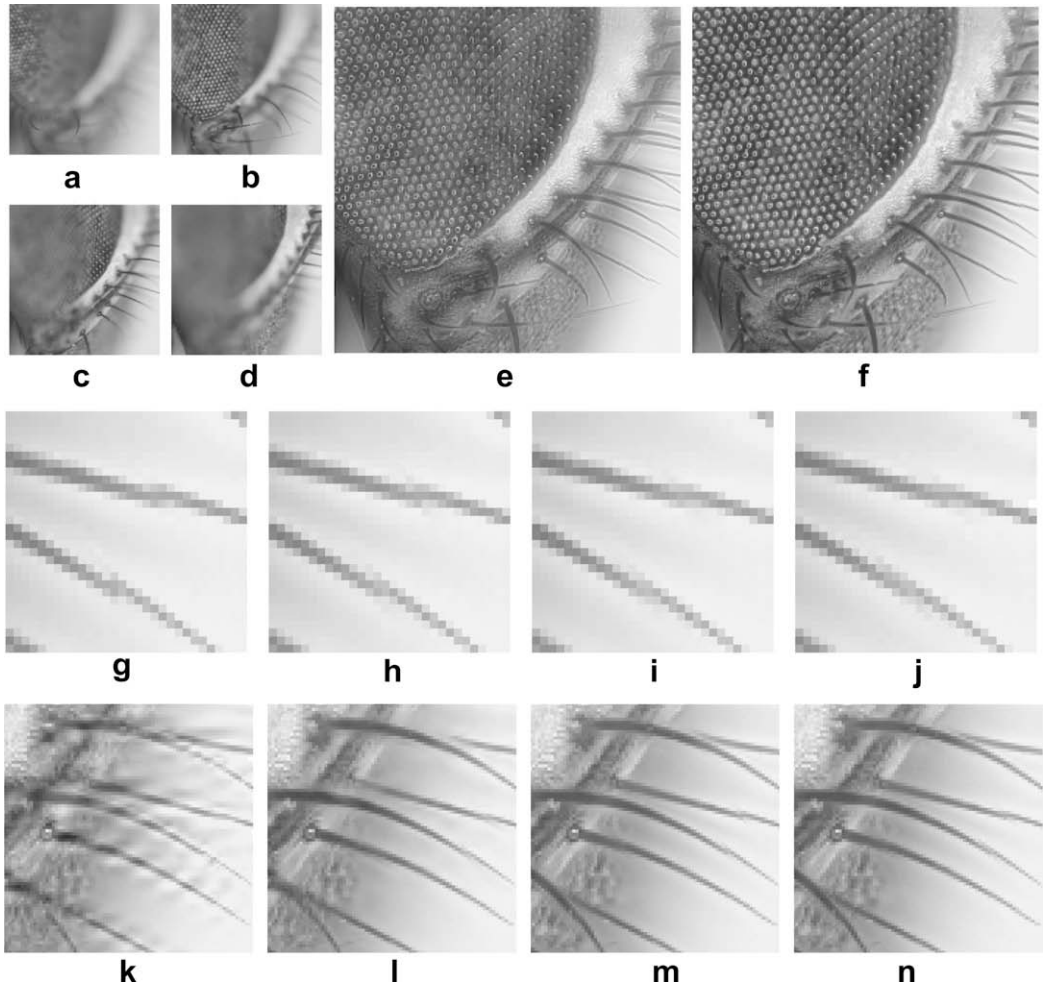

Fig. 13. Images of a real sample of a fly head captured from a bright-field microscope composed of 12 slices of size 236×236. (a–d) Four representative slices. (e, f) Fused image obtained through log-Gabor with decomposition levels $K = 3$ and $K = 4$, respectively, and the same number of orientations $P = 3$. (g–j) Magnified details of the fused images through log-Gabor with $P = 1, P = 3, P = 8$ and the Multisize Windows ($P = 3$), respectively. (k–n) Magnified details of the fused images obtained through decimated biorthogonal (bior4.4), biorthogonal (bior4.4), log-Gabor and log-Gabor + Multisize Windows, respectively ($P = 3$). All examples (g)–(n) were equally configured with $K = 4$ and $N = 7$, except for the decimated biorthogonal case with $N = 3$. The Multisize Windows used the window set and thresholds described in scenario 1.

Please cite this article in press as: R. Redondo et al., Multifocus image fusion using the log-Gabor transform and a Multisize Windows ..., Informat. Fusion (2008), doi:10.1016/j.inffus.2008.08.006

tative measures in simulated scenarios but also by visual inspection in a real microscopy experiment. It is obvious, however, that other types of scenarios would also help further investigate aspects of MR processes.

Frequency features placed at medium-frequency bands (the second and third scales) are beneficial in terms of overall blur elimination. This might be due to the fact that relevant/focused details are mainly present in the first octave of the Fourier domain. However, more than two or three levels of decomposition caused impairment of the decision maps. As the size of the window increases, the number of orientations diminishes its influence on decision maps. This could explain why the Laplacian pyramids achieve such remarkable performance. Nevertheless, more than three orientations improved delicate features in microscopy, but an elevated number of them produced blurring.

Decimation should be definitely avoided since it induces poor robustness and severe ringing artifacts. The visual quality supplied by all undecimated wavelets was notable. Log-Gabor, together with biorthogonal wavelets, yielded the least distorted decision maps, but in the presence of noise none of the tested wavelets was superior. In addition, the log-Gabor transform yielded fewer low-frequency artifacts around demanding structures, which could be ascribed to complex-valued coefficients. In future work, the impact of misregistered images (shifted and rotated) on fusion regarding shift-invariance properties of transforms could be investigated.

The MW technique deserves special attention since it delivered better decision maps than any of the other single-window approaches in the experiments, regardless of the decomposition level, noise level or type of wavelet. The sizes and shapes of the objects, the level of noise and the decomposition level condition the size of the averaging window. In real applications it is difficult to estimate the appropriate window size since there does not exist a ground truth image for comparison. Hence, the MW technique avoids the pre-fixation of the window size and estimation of the noise strength. Although a set of sizes has to be previously defined, here we simply covered a broad range of equidistant values, which can be taken as a general rule. Likewise, the set of confidence thresholds depends almost exclusively on the scale of the objects present in the channels, which can be automated in a straightforward way. Though the current MW setting performed quite stably, exhaustive explorations on the influence of the window and threshold sets may be of interest.

Lastly, the overall methodology was designed for multifocus frameworks but could be easily applied to other image fusion problems, such as multimodal fusion. Special attention should be paid to the combination of low-pass sub-bands. Use of match measures instead of the maximum selection rule could offer significant multisensor improvements but would put at risk compatibility with the MW technique.

Acknowledgements

Special thanks to the courtesy of Radim Sara for providing the microscopy images. RR is supported by a CSIC-I3P fellowship, FS is supported by the Spanish States Secretary of Education and Universities Fellowship Program and SF is supported by a MEC-FPU fellowship. This work has been additionally supported by projects TEC 2004-00834, TEC2005-24739-E, TEC2005-24046-E, PI040765, 2004CZ0009 CSIC-Academy of Sciences of the Czech Republic, No. 102/04/0155 and No. 202/05/0242 of the Grant Agency of the Czech Republic and No. 1M0572 (Research Center DAR) of the Czech Ministry of Education.

References

[1] Y. Wang, B. Lohmann, Multisensor image fusion: concept, method and applications, Tech. rep., Univ. Bremen., 2000.
[2] P.J. Burt, E.H. Adelson, The Laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[3] T. Pu, G. Ni, Contrast-based image fusion using the discrete wavelet transform, Opt. Eng. 39 (8) (2000) 2075–2082.
[4] P.J. Burt, A gradient pyramid basis for pattern-selective image fusion, Soc. Inform. Display Digest Tech. Papers 16 (1985) 467–470.
[5] A. Toet, A morphological pyramidal image decomposition, Pattern Recogn. Lett. 9 (4) (1989) 255–261.
[6] A. Toet, Image fusion by a ratio of low-pass pyramid, Patten Recogn. Lett. 9 (4) (1989) 245–253.
[7] H.H. Wang, A new multiwavelet-based approach to image fusion, J. Math. Imaging Vis. 21 (2) (2004) 177–192.
[8] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graph. Models Image Process. 57 (3) (1995) 235–245.
[9] A. Toet, Hierarchical image fusion, Mach. Vis. Appl. 3 (1) (1990) 1–11.
[10] Z. Zhang, R.S. Blum, A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application, Proc. IEEE 87 (8) (1999) 1315–1326.
[11] M. Subbarao, T. Choi, A. Nikzad, Focusing techniques, J. Opt. Eng. 32 (11) (1993) 2824–2836.
[12] M. Subbarao, J.K. Tyan, Selecting the optimal focus measure for autofocusing and depth-from-focus, IEEE Trans. PAMI 20 (8) (1998) 864–870.
[13] Y. Zhang, Y. Zhang, C. Wen, A new focus measure method using moments, Image Vis. Comput. 18 (12) (2000) 959–965.
[14] J. Kautsky, J. Flusser, B. Zitová, S. Šimberová, A new wavelet-based measure of image focus, Pattern Recogn. Lett. 23 (14) (2002) 1785–1794.
[15] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Inform. Fus. 4 (4) (2003) 259–280.
[16] G. Piella, H. Heijmans, A new quality metric for image fusion, in: IEEE International Conference on Image Processing, ICIP, 2003, pp. 173–176.
[17] A. Toet, E.M. Franken, Perceptual evaluation of different image fusion schemes, Displays 24 (1) (2003) 25–37.
[18] V. Petrović, C. Xydeas, Evaluation of image fusion performance with visible differences, in: Proceedings of ECCV, vol. 3023, 2004, pp. 380–391.
[19] S. Fischer, F. Sroubek, L. Perrinet, R. Redondo, G. Cristóbal, Self-invertible log-Gabor wavelets, Int. J. Comput. Vis. 75 (2) (2007) 231–246.
[20] S. Fischer, G. Cristóbal, R. Redondo, Sparse overcomplete Gabor wavelet representation based on local competitions, IEEE Trans. Image Process. 15 (2) (2006) 265–272.
[21] S. Fischer, R. Redondo, L. Perrinet, G. Cristóbal, Sparse approximation of images inspired from the functional architecture of the primary visual areas, EURASIP JASP, in press, 2007 (special issue on Visual Perception).
[22] R. Redondo, F. Sroubek, S. Fischer, G. Cristóbal, Multifocus fusion with multisize windows, in: A.G. Tescher (Ed.), Proceedings of the SPIE Applications of Digital Image Processing XXVIII, vol. 5909, 2005.
[23] J.M. Shapiro, Embedded image coding using zero trees of wavelet coefficients, IEEE Trans. Signal Process. 41 (12) (1993) 3445–3462.
[24] A. Srivastava, A.B. Lee, E.P. Simoncelli, S.C. Zhu, On advances in statistical modeling of natural images, J. Math. Imaging Vis. 18 (1) (2003) 17–33.
[25] D. Gabor, Theory of communication, J. Inst. Electr. Eng. 93 (1946) 429–457.
[26] W. Lawton, Applications of complex valued wavelet transforms to subband decomposition, IEEE Trans. Signal Process. 41 (12) (1993) 3566–3568.
[27] B. Forster, D. Van De Ville, J. Berent, D. Sage, M. Unser, Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images, Microsc. Res. Tech. 65 (1–2) (2004) 33–42.
[28] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multiscale transforms, IEEE Trans. Inform. Theory 38 (2) (1992) 587–607.
[29] I. Koren, A. Laine, F. Taylor, Image fusion using steerable dyadic wavelet transform, in: Proceedings of IEEE International Conference on Image Processing, vol. 3, 1995, pp. 232–235.
[30] O. Rockinger, Image sequence fusion using shift invariant wavelet transform, in: Proceedings of IEEE International Conference on Image Processing, vol. 3, 1997, pp. 288–291.
[31] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Prentice Hall, New Jersey, 2002.
[32] J. Portilla, V. Strela, M. Wainwright, E. Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain, IEEE Trans. Image Process. 12 (11) (2003) 1338–1351.
[33] R.J. Pieper, A. Korpel, Image processing for extended depth of field, Appl. Opt. 22 (10) (1983) 1449–1453.
[34] S.A. Sugimoto, Y. Ichioka, Digital composition of image with increased depth of focus considering depth information, Appl. Opt. 24 (14) (1985) 2076–2080.
[35] V. Tympel, New high-level image capture system for conventional microscopy, in: Proceedings of SPIE Medical Imaging: Image Display, vol. 2707, 1996, pp. 529–536.
[36] A.G. Valdecasas, D. Marshall, J.M. Becerra, J.J. Terrero, On the extended depth of focus algorithms for bright field microscopy, Micron 32 (2001) 559–569.
[37] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM J. Math. 29 (2) (1997) 511–546.