Region-based multifocus image fusion for the precise acquisition of Pap smear images

Santiago Tello-Mijares
Jesús Bescós

Region-based multifocus image fusion for the precise acquisition of Pap smear images

Santiago Tello-Mijares[a,b,*] **and Jesús Bescós**[a]
[a]Universidad Autónoma de Madrid, Escuela Politécnica Superior, Video Processing and Understanding Lab, Madrid, Spain
[b]Instituto Tecnológico Superior de Lerdo, Department of Postgraduate, Lerdo, Mexico

Abstract. A multifocus image fusion method to obtain a single focused image from a sequence of microscopic high-magnification Papanicolau source (Pap smear) images is presented. These images, captured each in a different position of the microscope lens, frequently show partially focused cells or parts of cells, which makes them unpractical for the direct application of image analysis techniques. The proposed method obtains a focused image with a high preservation of original pixels information while achieving a negligible visibility of the fusion artifacts. The method starts by identifying the best-focused image of the sequence; then, it performs a mean-shift segmentation over this image; the focus level of the segmented regions is evaluated in all the images of the sequence, and best-focused regions are merged in a single combined image; finally, this image is processed with an adaptive artifact removal process. The combination of a region-oriented approach, instead of block-based approaches, and a minimum modification of the value of focused pixels in the original images achieve a highly contrasted image with no visible artifacts, which makes this method especially convenient for the medical imaging domain. The proposed method is compared with several state-of-the-art alternatives over a representative dataset. The experimental results show that our proposal obtains the best and more stable quality indicators. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JBO.23.5.056005]

1 Introduction

Methods used in microscopic autofocus systems (MAFSs) are generally based on maximizing a focus measure obtained from a captured image; such measure is evaluated as a function of the lens' Z axis position. There are plenty of works that report algorithms to control the motion of the lens along the Z axis in order to efficiently find the best position: based on the evaluation of that focus measure, they obtain the best focused single image.[1]

In the case of a MAFS applied to cytology observation, when working with targets of high magnification—around 40×—one of the principal issues is that in many cases, cells are in fact located on different Z levels in the slide, even corresponding to different ranges of the depth-of-field (DoF) of the lens. In these cases, the best focused image selected by a classical autofocus method will include unfocused parts, hence missing important information. A solution is to somehow combine a set of images captured with different DoFs to obtain a fully focused image.

Image fusion is the process of combining relevant visual information (i.e., important, complementary, and redundant) from multiple input images into a single resulting image. This should be achieved without introducing artifacts, in a way in which the resulting image contains more accurate, stable, and complete information[2–4] than the input images, therefore making it more suitable for human perception and for later processing operations (e.g., segmentation and feature extraction). For the

autofocus application, this resulting image might be obtained by applying multifocus image fusion (MFIF) techniques: a set of N images is captured from a static scene at different focus levels, and the focused objects in this set of images are fused together to create a sharp image with all those relevant objects fully focused.[1,3–5] While the reported MFIF methods are generally applied to fuse two images of a scene, the underlying techniques can be adapted to fuse a larger set of images, as we propose.

The context of our work is a project (see ACK section for details) to help early diagnosis of cervical intraepithelial neoplasia in the rural areas of the Coahuila State (Mexico). In these areas, for cultural reasons, the refusal among women to go to the capital persists until the symptoms of the disease are unbearable. The objective of the project is to enable the Papanicolau test, facilitating through its automation the taking of tissue samples in the rural area and the telematic sending of selected images of these samples for diagnosis by specialists. This requires capturing hundredths of focused images per tissue sample (see our autofocus contributions in Ref. 6) and analyzing of these images to identify and segment cervical nuclei (see our contributions on this area in Ref. 7). In this paper, we target the enhancement of the captured images via MFIF techniques.

MFIF techniques operate on a set of N input images of a single scene, each with a different DoF. Overall, these images are first partitioned into generally homologous regions. Regions sharing a same location in the set of images are then evaluated, and the region with the highest focus measure is selected; finally, selected regions, usually from different images, are fused to compose the final focused image. There are many MFIF

*Address all correspondence to: Santiago Tello-Mijares, E-mail: jtello@itslerdo.edu.mx

algorithms reported in the literature; a comprehensive review can be found in Refs. 8 and 9. The MFIF technique is broadly used in many application fields, such as microscopy,[10] biology,[11] and medical imaging.[12]

In this paper, we analyze and compare with works using focus measures like those we used in our base autofocus system (i.e., transformed-domain measures[6]), for a fair comparison. In this direction, the work in Ref. 1 applies an 8×8 block discrete cosine transform (DCT) to two input images; then, it compares homologous blocks of both images using the variance of the coefficients and selects the block with the highest value; finally, it applies the inverse DCT to the image composed by the selected blocks. They also propose a variant which applies a consistency verification index for block selection, in order to enhance the resulting image quality. The work in Ref. 4 is similar to that in Ref. 1 but using a different measure to compare blocks. Methods based on blocks generally present artifacts, because parts of the focused cells in different levels of the DoF might belong to a same block. Kumar[5] uses the discrete cosine harmonic wavelet transform (DCHWT), a multiscale technique (three levels, in this case). As these multiscale methods involve decimation, most pixels in the resulting image do not keep original pixel values of any of the source images. Recently, some fusion techniques operating in the gradient domain have been reported, as that in Ref. 3. This work, also using a multiscale approach, uses a focus measure based on the saliency structure of the image, and it is designed to operate on well-known images (i.e., flower, clock, pepsi, etc.) with just two objects (one focused and the other unfocused).

Some recent works present similar fusion techniques applied to combine different sources of information into a single image but not necessarily due to a multifocus situation. Liu et al.[13] described a MFIF method that separates source images into "cartoon content" and "texture content" via an improved iterative reweighted decomposition algorithm; fusion rules are designed to separately fuse both types of content, and finally, the fused cartoon and texture components are combined. The technique naturally approximates the morphological structure of the scene. The work in Ref. 14 presents a medical application to diagnose vascular diseases; they use a type of wavelet transform combined with an averaging-based fusion model to fuse osseous and vascular information together; they present a rapid MFIF algorithm, less complex but still very effective, with very low memory requirements. For a similar objective, Dogra et al.[15] propose an effective image fusion method also working on the wavelet domain, along with a preprocessing of the source images with a selected sequence of spatial and transformed-domain techniques to create a highly informative fused image for osseous-vascular 2-D data.

In this paper, we propose an MFIF method that analyzes sequences of up to 15 microscopy input images corresponding to different levels of DoF of a same "slide-scene." We propose (Sec. 2) an object-based approach, which dramatically reduces the visibility of fusion-generated artifacts while keeping focused parts of the input images intact. To evaluate our results, we compare with five different existing techniques (Sec. 3) by testing over 50 realistic and practical Pap smear sequences of images, and over the two blurred microscopic images provided by Ref. 11. Finally, conclusions are presented in Sec. 4.

2 Proposed Multifocus Microscope-Image Fusion Method

Figure 1 illustrates the general flow of the proposed MFIF method, which is further detailed in the following subsections. The starting point is a set of images (15 in our experiments, but 2 in other works we compare to) captured with the lens in a varying position of the Z axis. The first step, which is not the topic of this paper, is the selection of the "best-focused image" of the set (see details in Ref. 6). Let $\{I_i, i = 1 \ldots N\}$ be this set of input images (Fig. 1) and let I_{bf} be the best-focused image, being $i = bf$ its index in the set. This image is first coarsely segmented to identify its main regions or objects, which are considered the main scene objects, each represented by a binary mask. Then, for each scene object or segmented region, its mask is applied to the set of input images, and a focus measure is obtained for that region in every image of the set. A preliminary image, which we name "combined image, I_c," is then generated
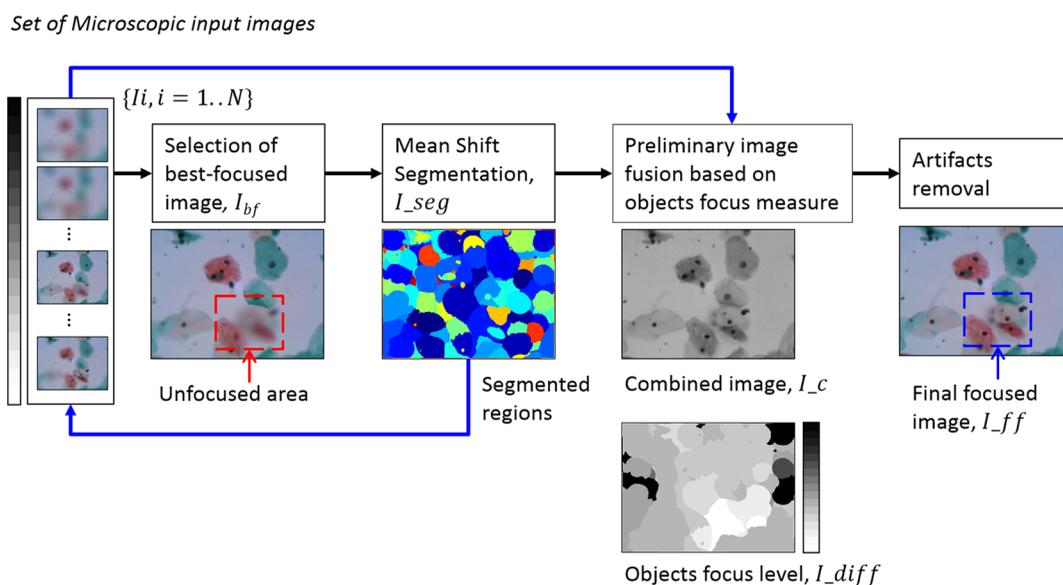
Fig. 1 Proposed multifocus image fusion method for a set of microscopic high-magnification Pap smear images.

by replacing in the best-focused image each segmented region with the corresponding best-focused region in the set. Finally, the removal of the artifacts generated in the contours of these combined regions is performed by a total variational-based filter[16] to obtain the "final focused image, I_ff."

2.1 Mean-Shift Segmentation

We use the mean-shift algorithm[17] to obtain a coarse segmentation, I_seg, of the best-focused image, I_{bf} (Fig. 1) into N_c regions or clusters. Mean-shift is a nonparametric technique for analyzing multimodal data that has multiple applications in pattern analysis,[18] including its use for image segmentation. We start from the observation that cells have a predetermined size and colors that are always much darker than the background. We characterize each image pixel by a vector $[L, a, b, x, y]$ or $[L, x, y]$, depending on whether the input images are RGB or gray: $[L, a, b]$ describes the pixel color, and $[x, y]$ its coordinates. We then run the mean-shift algorithm over this five-dimensional or three-dimensional distribution with a bandwidth value $h = 20$, which was selected so that cell regions and background are segmented in more than one cluster; this is required for the next assigning process to be effective. A proper selection of the h parameter is somehow application dependent: it should be larger than the smallest nonfocused region. However, if this requirement is met, its effect on the results is negligible. Consider that block-based approaches also prefer unfocused regions to be greater than the block size, but there is usually no flexibility in the selection of this size.

2.2 Preliminary Image Fusion Based on the DCT Focus Measure

The next step is to generate a "combined image" (follow this process in Fig. 2), I_c, which is a merging of the best focused parts of the set of input images. First, a focus measure is locally obtained for every image of the set, following the method described in Ref. 6: in brief, for every image of the input set, I_i, a 8×8 block DCT is performed, the sum of the absolute value of its 32 lower-frequency AC coefficients is calculated, and a same-size energy image, E_i, is obtained by assigning each pixel the calculated energy of its corresponding block. This results in a set of DCT energy images, $\{E_i, i = 1 \ldots N\}$ [Fig. 2(a)].

The topology of the I_c image is the same of that of the segmented image, I_seg. Every cluster or region in I_seg is used to

generate a mask, $\{M_r, r = 1 \ldots N_c\}$ [Fig. 2(b)]. For every region, its corresponding mask is applied over every energy image, $\{E_i, i = 1 \ldots N\}$, and the mean energy of the masked region is calculated for every such energy image. The index, $i\_$max, of the energy image showing maximum energy for that region is obtained; then, the corresponding region of the I_c image is initialized with the pixel values of the homologous region of the I_{i_max} image from the $\{I_i, i = 1 \ldots N\}$ set. In parallel, in the I_diff image (see Fig. 2), we keep for every region the absolute difference between the $i\_$max index and the $i = $ bf index of the best-focused image, which somehow indicates the degree of out-of-focus of such region, or the object focus level, ranging from 0 (black-level: the region is best-focused in the I_{bf} image) to N (white-level: the region is best-focused in the image with the worst global focus measure).

As opposed to other methods, such as Refs. 1 and 4, where the local focus comparison among the set of input images is performed block by block, we propose to compare region by region using the segmentation of the best-focused image to define such regions. This avoids highly visible block artifacts appear anywhere. Instead, contour artifacts might appear in the boundaries of the identified regions, being here much less visible. The visibility of these contour artifacts depends on the aforementioned degree of out-of-focus of each combined region. In the next subsection, we propose to use a total variational-based filter to eliminate the contour artifacts of the combined image, I_c.

2.3 Artifacts Removal

The next step is to generate the final focused image, I_ff, by attenuating the artifacts or false contours that may appear in the combined image, I_c, due to merging regions from different input images. We propose to attenuate artifacts by applying a total variational-based diffusion method.[16] This method will only be applied in the artifacts-prone areas, according to the information in the I_diff image, hence preserving or keeping intact most of the image pixels. Observe that the method aims to mitigate these false contours, not real object contours.

2.3.1 Generation of a mask of the artifact-prone areas

In Fig. 3, we show several examples of artifacts generated at the boundaries of the regions of three I_c images. Our proposal is to process I_c pixels only at the edges defined by I_diff, i.e., only at the boundaries between regions with different degree of focus,
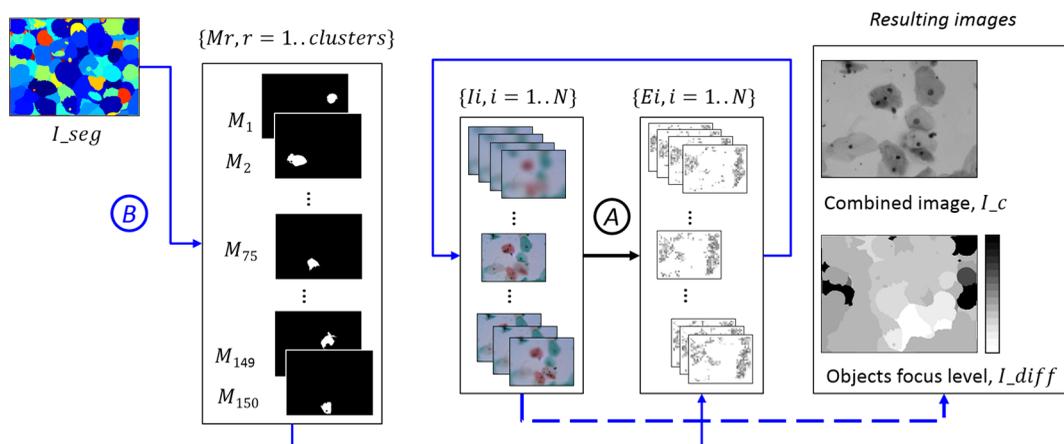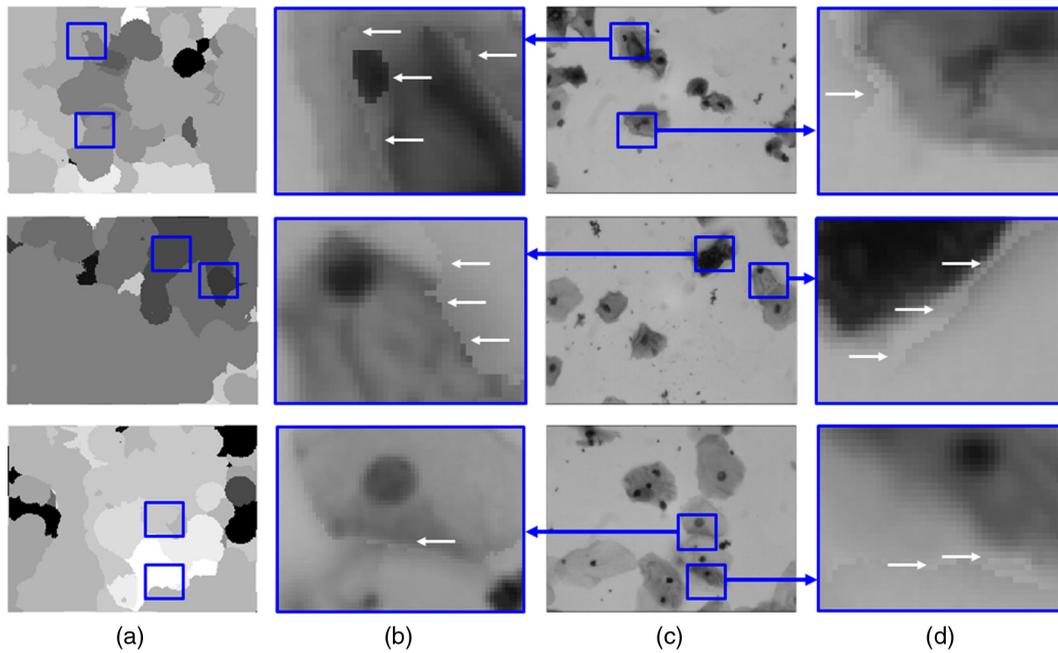

Fig. 2 Obtaining the combined image, *I_c* form the set of input images, $\{I_i, i = 1, \ldots, N\}$.

Fig. 3 Examples of contour artifacts in the boundaries of the merged regions: (a) *I_diff*; (c) *I_c*; (b) and (d) detailed images of the artifacts (see white arrows) generated by the regions merging.

in order to obtain an image without artifacts on these boundaries while keeping original pixels in most of the resulting image. For this purpose, we first obtain an edges image from I_diff (see Fig. 4). Then, as the extent of the artifacts between adjacent regions is expected to be proportional to the difference between their degrees of focus, we perform an adaptive morphological dilation over the thresholded edges image, using a structural element with a size proportional to the intensity of every edge. The resulting mask, M_artifacts (see Fig. 4), will define where the following enhancement steps will be applied.

A main contribution of our method is that artifacts removal is only performed in the areas that may include them, hence preserving original pixels in most of the image, which is critical for medical imaging applications. Works in Refs. 1 and 4, as they perform image fusion over DCT blocks, are prone to generate block-artifacts, which are not later eliminated. In the multiscale methods, such as Refs. 5 and 3, the original pixels are not usually preserved in the fused image: the resulting image in Ref. 5 does not present artifacts due to the nature of the method, which modifies pixels intensity via averaging, resulting in a smoother image; the method proposed in Ref. 3 eliminates artifacts just in the "unknown zone," which is a predefined area in the boundary generated between the two considered source images with two different focus levels.

2.3.2 *Artifacts removal via total-variation filtering*

Let us consider that the combined image I_c, which contains contour artifacts in the contours defined by the M_artifacts mask, is a noisy image; let I_ff, the final focused image, be the desired sharp and clean image. We can then declare that I_c = I_ff + n, where n is the aditive noise, which we assume concentrated in the pixels indicated by M_artifacts. We obtain I_ff from I_c using a total variation filter. These filters were first suggested by Ref. 16 and are based on the minimization of an energy functional, subject to the constraint $u - u_0 = \sigma^2$, where u, u_0, and σ^2, for this work, are, respectively, the gradient of the true image (I_ff), the gradient of observed image (I_c), and the variance of the noise $n = (u - u_0)$. Then, the iterative equation to obtain the desired clean image is I_ff^t = I_c$^t - \lambda(u - u_0^t)$, where I_c^0 = I_c, and λ is a regularization parameter which we set to $\lambda = 0.1$ to preserve the smallest structures.

In order to obtain u_0 for the first iteration, we apply a Laplacian filter to I_c (Fig. 5). To estimate u, we consider that its gradient equals that of I_c except for the edge areas defined by M_artifacts. In these areas, for every pixel, we assume that u equals the gradient, $\{u_i, i = 1 \ldots N\}$, of the source image showing maximum local variance around such pixel. The gradients of the source images are also obtained
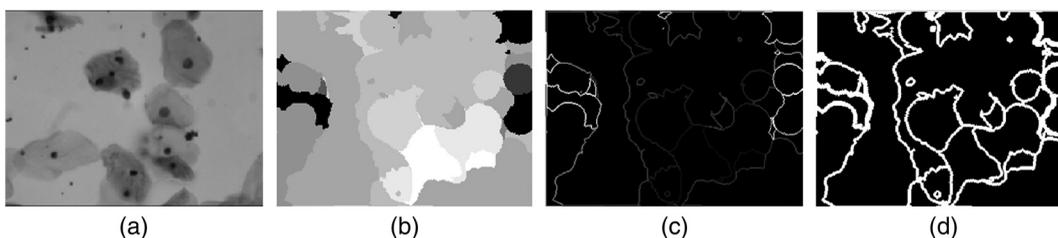
Fig. 4 Example of the generation of the artifacts mask: (a) *I_c*, (b) *I_diff*, (c) edges in *I_diff*, and (d) dilated and thresholded edges in *M_artifacts*.

(a)

(b)
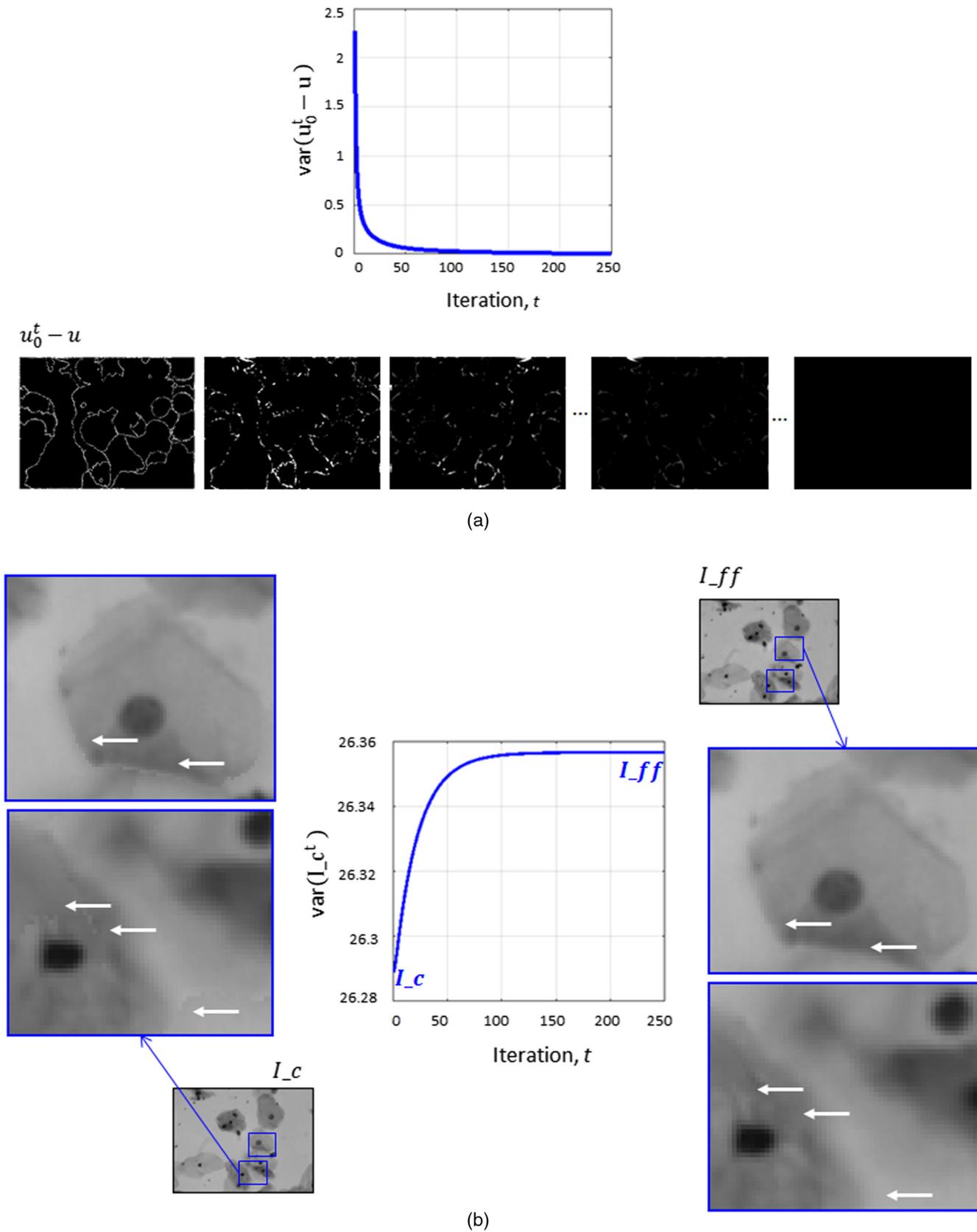
Fig. 5 Example of the iterative artifacts elimination process: (a) evolution of the gradients difference $u_0^t - u$ and (b) comparison of the I_c image and of the I_ff image obtained after the process converges; white arrows indicate the main removed artifacts.

applying a Laplacian filter, and the local variance is computed in 3×3 windows. Once we get I_ff^t for this first iteration, i.e., I_ff^0, we set I_c^{t+1} = I_ff^t and repeat the process until it converges to I_ff. Figure 5 shows an example of the evolution of the variance of the gradient difference, $u - u_0^t$, and of the obtained image, I_ff^t, for every iteration.
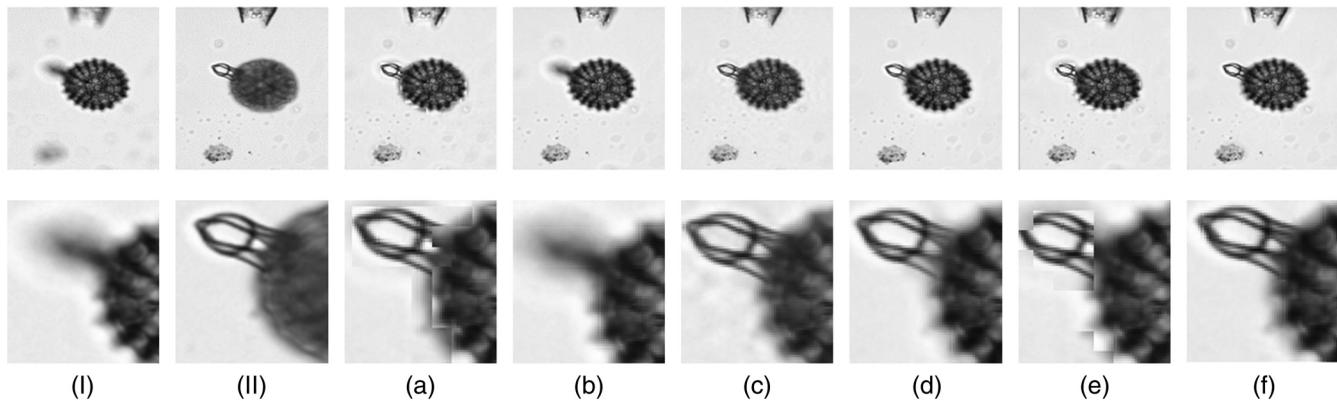
3 Experimental Results

To assess the potential of our approach, we compare the proposed method with the works reported in Refs. 1-1, 1-2, 5, 3 and 4, as we can see in Figs. 6–8 captions, and Tables 1 and 2. The code to run these reported algorithms was kindly provided by every author: the software is available together with the papers. The set of images used for the experimental evaluation, hereinafter the dataset, consist of the microscopic image pair from the MFIF reported in Ref. 11 (see Fig. 6) and a set of 50 Pap smear image sequences (320×240 pixels in RGB),

each containing 15 images with different DoF and focused cells in several of them (see examples in Figs. 7 and 8).

3.1 Quality Metrics

The objective evaluation of a fused image is a difficult task because there is no universally accepted metric to evaluate an image fusion process.[2] A frequent solution is the use of different metrics to test the fusion results from different viewpoints.[19] Quality metrics for MFIF can be classified depending on the availability of the target image:[20] metrics known as full-reference assume that a complete reference image (distortion-free) is available; however, in many practical applications, the reference image is not available; so, "no-reference" or "blind" quality metrics are used. As our dataset includes images captured in practical situations, we do not account for reference images. An alternative to these MFIF-based quality metrics is to evaluate focus metrics on the resulting fused image, as the aim in this

Fig. 6 Visual results for the first experiment. The top row shows full images and the bottom one shows the corresponding image in detail. Columns include source images A (I) and B (II); resulting images for the compared methods 1-1 (a) Haghighat, 2011(1), 1-2 (b) Haghighat, 2011(2), 5 (c) Kumar, 2013, 3 (d) Zhou, 2014 and 4 (e) Phamila, 2014; (f) the resulting image, I_ff.
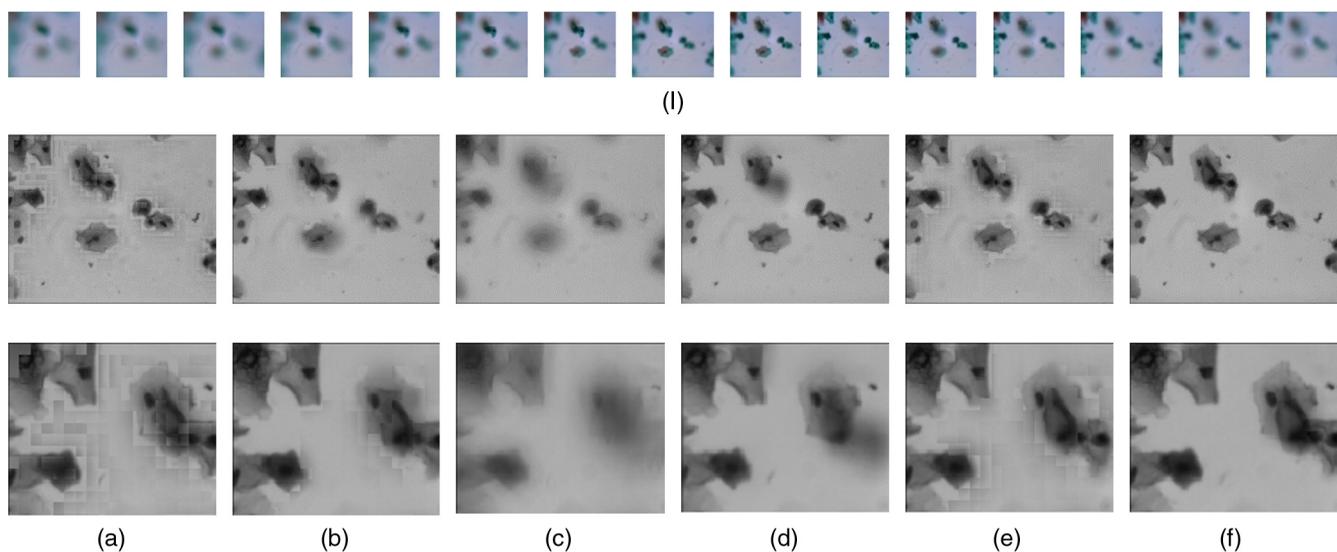
Fig. 7 Visual results for the second experiment. The top row shows the sequence of input images for sequence 3 (I). The bottom two rows show full images and the corresponding image detail for the final fused images obtained by each compared method: 1-1 (a) Haghighat, 2011(1), 1-2 (b) Haghighat, 2011 (2), 5 (c) Kumar, 2013, 3 (d) Zhou, 2014, 4 (e) Phamila, 2014 and (f) the resulting image, I_ff.
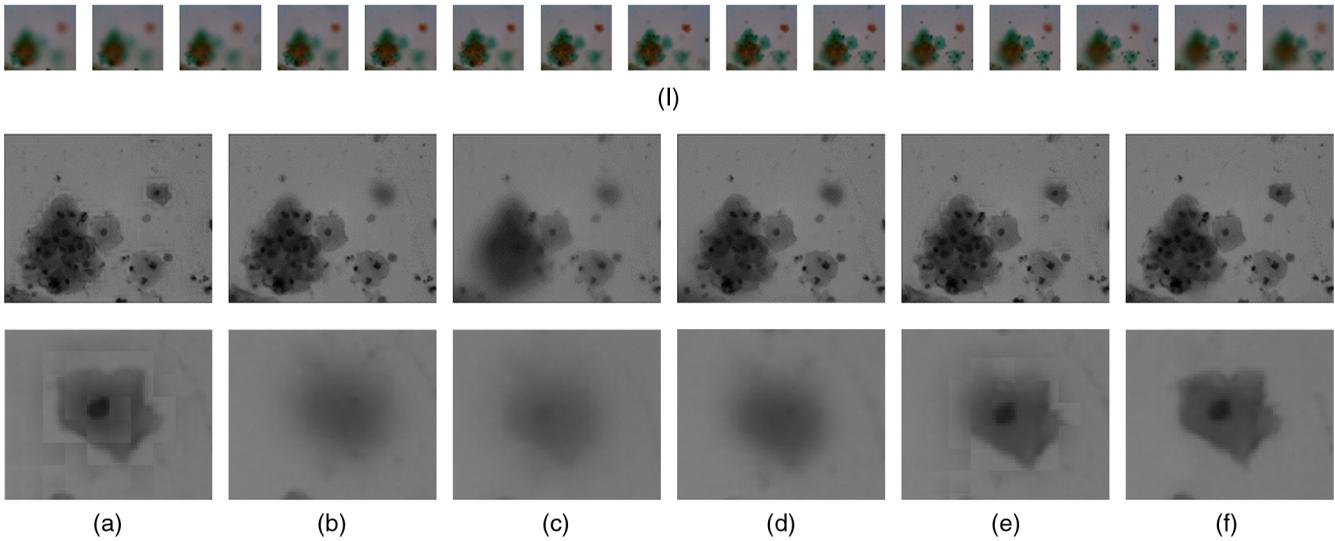
Fig. 8 Visual results for the second experiment. The top row shows the sequence of input images for sequence 37 (I). The bottom two rows show full images and the corresponding image detail for the final fused images obtained by each compared method: 1-1 (a) Haghighat, 2011(1), 1-2 (b) Haghighat, 2011 (2), 5 (c) Kumar, 2013, 3 (d) Zhou, 2014, 4 (e) Phamila, 2014 and (f) the resulting image, I_ff.

scenario is to obtain a perfectly focused image. We describe below the metrics we have used.

No-reference metrics—The Petrovic metrics,[21,22] based on gradient information, include three indicators: $Q^{AB/F}$, which represents in a normalized way the total information transferred from the source images (A, B) to the fused image (F); and $L^{AB/F}$ and $N^{AB/F}$, which evaluate the complement to $Q^{AB/F}$, i.e., the loss of information, but just considering locations, where the gradient of the source images is greater $(L^{AB/F})$ or lower $(N^{AB/F})$ than that of the fused image. We have computed the $Q^{AB/F}$ indicator as an overall quantitative measure of the fusion quality. For $M \times N$ images, $Q^{AB/F}$ is obtained according to

$$Q^{AB/F} = \sum_{\forall n,m}^{N,M} Q_{n,m}^{AF} w_{n,m}^{A} + Q_{n,m}^{BF} w_{n,m}^{B} / \sum_{\forall n,m}^{N,M} w_{n,m}^{A} + w_{n,m}^{B}, \quad (1)$$

where Q^{AF} and Q^{BF} estimate edges preservation from the A and B source images, and w^A and w^B are local perceptual weighting factors usually corresponding to the gradients of these source images. A value of $Q^{AB/F} = 0$ means complete loss of information and $Q^{AB/F} = 1$ represents ideal fusion. This indicator is defined for the case of two source images (A and B); we have adapted it to the dataset sequences including 15 source images (from I_1 to I_{15}):

$$Q = Q^{I_1 \cdots I_{15}/F} = \frac{\sum_{\forall n,m}^{N,M} Q_{n,m}^{I_1 F} w_{n,m}^{I_1} + \ldots + Q_{n,m}^{I_{15} F} w_{n,m}^{I_{15}}}{\sum_{\forall n,m}^{N,M} w_{n,m}^{I_1} + \ldots + w_{n,m}^{I_{15}}}.$$

$$(2)$$

Focus metrics—These include the standard deviation of the normalized pixel intensities of the fused image, F, the entropy of this image, the average gradient magnitude, which indicates sharpness, etc. We have selected the standard deviation, because it has been demonstrated to provide the best overall performance in estimating the focus level for nonfluorescence microscopy applications, including Pap smear and blood smear samples.[23–26]

3.2 Experiments and Discussion

The first experiment is conducted over two microscopic images kindly provided by Ref. 11, each showing different focused parts of the same object [see Figs. 6(I) and 6(II)]. We have applied to these source images the aforementioned five fusion algorithms and our proposed method. Figure 6 shows the resulting images along with a detail of each, in order to visually or qualitatively assess the performance of each method. Table 1 includes data with the quantitative evaluation of this first experiment.

From a qualitative point of view, we observe that methods Refs. 1-1 and 4 [Figs. 6(a) and 6(e)] present highly visible block artifacts; these methods compare the DCT energy in homologous 8×8 blocks, which generates comparison errors when the images contain nonsquare elements in different depths of field or when cervical cells are round. The enhancement proposed by Ref. 1-2 [Fig. 6(b)], based on a consistency verification index to decide which block is selected, removes block artifacts in this example but at the expense of a poor visual result. The

Table 1 Quantitative results for the first experiment: performance quality metrics for the final fused microscope image obtained by each method.

	1-1 Fig. 6 (a) Haghighat, 2011(1)	1-2 Fig. 6 (b) Haghighat, 2011(2)	5 Fig. 6 (c) Kumar, 2013	3 Fig. 6 (d) Zhou, 2014	4 Fig. 6 (e) Phamila, 2014	I_ff Fig. 6 (f)
SD	0.9877	0.9783	0.9534	0.9957	0.9843	1
$Q^{AB/F}$	0.8605	0.7629	0.8829	0.8884	0.8436	0.8919

multiscale approach proposed in Ref. 5 [Fig. 6(c)], which does not keep original pixel values, presents a noticeable contrast reduction. The method reported in Ref. 3 and the proposed method [Figs. 6(d) and 6(f)] yield acceptable visual results.

From a quantitative point of view (see Table 1), an interesting observation is to contrast the correlation between each measure of quality and the perceived visual result: the SD measure yields very good values for images with highly noticeable block artifacts [in case of Figs. 6(a) and 6(e)], because these artifacts increase image variance; the $Q^{AB/F}$ measure seems to be more in line with the qualitative findings.

Independently of these observations, Table 1 indicates that the proposed method behaves better in the light of both quality measures.

The second experiment targets the 50 sequences of Pap smear images obtained from the autofocus operation of a microscope. While reported works have focused on fusing two blurred images, many of them have applied their method in an iterative way to more than two input images, which is our practical context. Figures 7 and 8 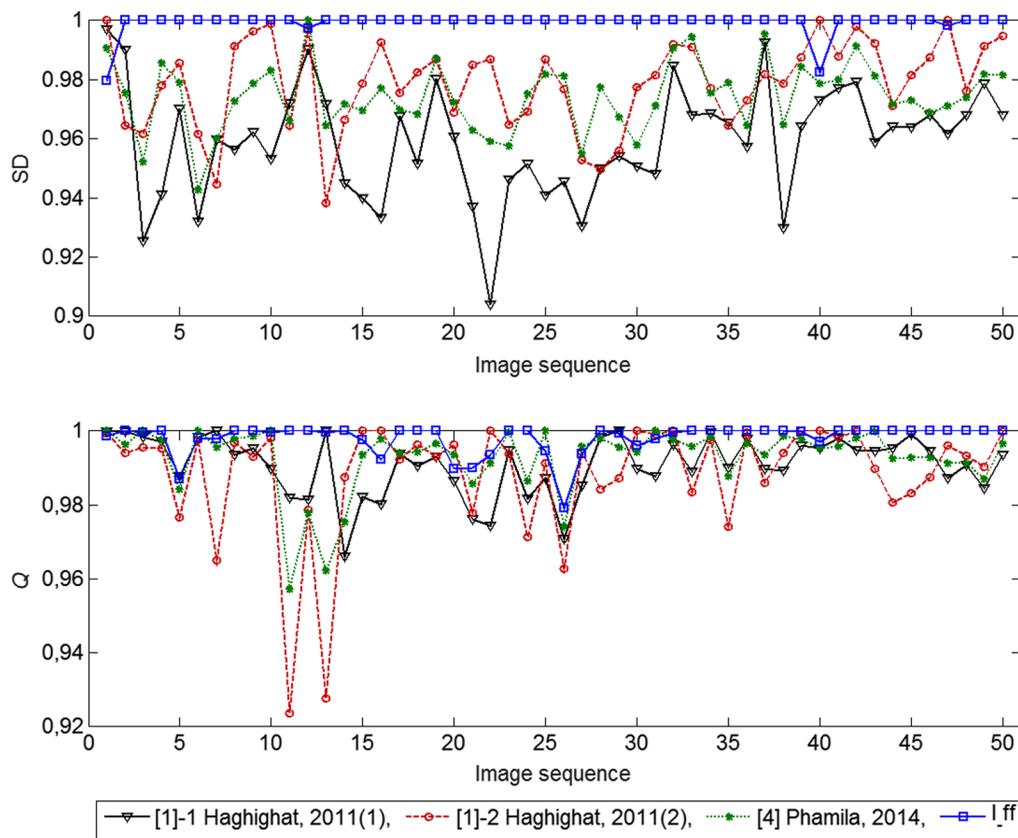show qualitative results for two of these sequences, and Table 2 and Fig. 9 compile the quantitative evaluation for the 50 sequences.

From a qualitative point of view, we clearly observe in Fig. 7 that the methods based on 8×8 DCT blocks [Figs. 7(a), 7(b) and 7(e), 8(a), and 8(e)] cannot avoid generating block artifacts. We can also observe that the multiscale approaches [Figs. 7(c), 7(d), 8(c), and 8(d)], which somehow process original pixels so that their value is never directly transferred to the final image, suffer from a severe loss of definition when the technique is applied to a large number of source image (15 images, instead of 2, for this experiment): several of the objects of interest are averaged, resulting in a loss of information and even the loss of complete cells. This is the situation for the method proposed in Ref. 3 [Figs. 7(d) and 8(d)], which, while losing very few information and objects of interest, sometimes loses full objects because it only compares two areas or regions in the image (focused and unfocused).

From a quantitative point of view, Table 2 indicates that the proposed method also behaves better for this part of the dataset including 50 image sequences. Apart from the mean values of

Table 2 Quantitative results for the second experiment: performance quality metrics (mean and deviation) for the final fused microscope images obtained by each method applied to the 50 image sequences.

	1-1 Haghighat, 2011(1)	1-2 Haghighat, 2011(2)	5 Kumar, 2013	3 Zhou, 2014	4 Phamila, 2014	I_ff
SD	0.9632 ± 0.0187	0.9832 ± 0.0134	0.8397 ± 0.0726	0.9424 ± 0.0359	0.9785 ± 0.0106	0.9987 ± 0.0045
Q	0.9920 ± 0.0072	0.9897 ± 0.0152	0.8941 ± 0.1011	0.6630 ± 0.0678	0.9940 ± 0.0081	0.9974 ± 0.0041

Fig. 9 Quantitative results for the second experiment. Performance quality metrics SD and Q for the final fused microscope images obtained by each method for every image sequence.

the quality indicators, Table 2 includes their standard deviation, which proves that the results obtained by the proposed method are also the most stable. Finally, Fig. 9 intends to further illustrate the stability of the tested methods that obtained better global results. We observe that the proposed method systematically outperforms other approaches in the light of these quality indicators.

4 Conclusion

This paper presents an object-oriented approach to the problem of obtaining a single focused image from a set of microscopic images captured from a single slide including objects that happen to be focused each in a different image of the set. The proposed MFIF method shows several specific advantages respect to other state-of-the-art methods: first, it is driven by a region-based segmentation, which prevents for the highly visible artifacts that may appear in block-based methods; second, it does not apply any kind of image transform, hence respecting the pixel-values of all focused regions, which is crucial for medical imaging applications; and finally, it includes a artifacts-removal technique, which only operates were required and adapts to the expected extent of the fusion-generated artifacts. Results, obtained over a representative dataset and compared to other published approaches, prove the validity of our proposal.

Appendix A: Pap Smear Images Sequences Dataset for Multifocus Image Fusion

A.1 Extra Material for Download

The extra materials are available for download (Ref. 27) and contain the following: the entire 50 cervical cells images sequences dataset; the region-based MFIF method proposed for comparison with the other MFIF methods (as a MATLAB interface); and the entire fusion results.

Disclosure

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

1. M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Comput. Electr. Eng.* 37(5), 789–797 (2011).
2. H. Zhao et al., "Multi-focus image fusion based on the neighbor distance," *Pattern Recognit.* 46(3), 1002–1011 (2013).
3. Z. Zhou, S. Li, and B. Wang, "Multi-scale weighted gradient-based fusion for multi-focus image," *Inf. Fusion* 20, 60–72 (2014).
4. Y. Phamila and R. Amutha, "Discrete cosine transform based fusion of multi-focus images for visual sensor networks," *Signal Process.* 95, 161–170 (2014).
5. B. S. Kumar, "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform," *Signal Image Video Process.* 7(6), 1125–1143 (2013).
6. S. Tello-Mijares et al., "Efficient autofocus method for sequential automatic capturing of high-magnification microscopic images," *Chin. Opt. Lett.* 11(12), 121102 (2013).
7. S. Tello-Mijares, J. Bescós, and F. Flores, "Nuclei segmentation and identification in practical pap-smear images with multiple overlapping cells," *J. Med. Imaging Health Inf.* 6(4), 992–1000 (2016).
8. L. Chen, J. Li, and C. L. Chen, "Regional multifocus image fusion using sparse representation," *Opt. Express* 21(4), 5182–5197 (2013).
9. Z. Liu et al., "Fusing synergistic information from multi-sensor images: an overview from implementation to performance assessment," *Inf. Fusion* 42, 127–145 (2018).
10. L. Kong et al., "Multifocus confocal Raman microspectroscopy for rapid single-particle analysis," *J. Biomed. Opt.* 16(12), 120503 (2011).
11. J. Tian and L. Chen, "Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure," *Signal Process.* 92(9), 2137–2146 (2012).
12. Q. Guihong, Z. Dali, and Y. Pingfan, "Medical image fusion by wavelet transform modulus maxima," *Opt. Express* 9(4), 184–190 (2001).
13. Z. Liu et al., "A novel multi-focus image fusion approach based on image decomposition," *Inf. Fusion* 35, 102–116 (2017).
14. A. Dogra, B. Goyal, and S. Agrawal, "Bone vessel image fusion via generalized reisz wavelet transform using averaging fusion rule," *J. Comput. Sci.* 21, 371–378 (2017).
15. A. Dogra et al., "Efficient fusion of osseous and vascular details in wavelet domain," *Pattern Recognit. Lett.* 94, 189–193 (2017).
16. L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D* 60(1), 259–268 (1992).
17. K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory* 21, 32–40 (1975).
18. D. Comaniciu and P. Meer, "Mean-shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 603–619 (2002).
19. X. Xia, S. Fang, and Y. Xiao, "High resolution image fusion algorithm based on multi-focused region extraction," *Pattern Recognit. Lett.* 45, 115–120 (2014).
20. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* 13(4), 600–612 (2004).
21. C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.* 36(4), 308–309 (2000).
22. V. Petrovic and C. Xydeas, "Objective image fusion performance characterization," in *Tenth IEEE Int. Conf. on in Computer Vision (ICCV '05)*, Vol. 2, pp. 1866–1871 (2005).
23. A. Santos et al., "Evaluation of autofocus functions in molecular cytogenetic analysis," *J. Microsc.* 188(3), 264–272 (1997).
24. Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: selecting the optimal focus algorithm," *Microsc. Res. Tech.* 65(3), 139–149 (2004).
25. X. Y. Liu, W. H. Wang, and Y. Sun, "Autofocusing for automated microscopic evaluation of blood smear and pap smear," in *28th Annual Int. Conf. of the IEEE In Engineering in Medicine and Biology Society (EMBS '06)*, pp. 4718–4721 (2006).
26. X. Y. Liu, W. H. Wang, and Y. Sun, "Dynamic evaluation of autofocusing for automated microscopic analysis of blood smear and pap smear," *J. Microsc.* 227(1), 15–23 (2007).
27. S. Tello-Mijares, "Multi focus image fusion," 2018, https://drive.google.com/drive/folders/1bcrJM8Kw0mzKF8VKpHg1WVlOxWGulwdj.

Santiago Tello-Mijares Received his BS degree in electronic engineering in 2006 and his PhD degree in electrical engineering science in 2013, from Instituto Tecnológico de la Laguna, Torreón, México; and in 2017, the PhD degree in telecommunications and informatics engineering at Universidad Autonóma de Madrid, Madrid, Spain. He is actually titular professor at Postgraduate Department in Instituto Tecnológico Superior de Lerdo, Lerdo, Mexico. His research interests are biomedical image, artificial intelligence, and robotics.

Jesús Bescós received his BS degree in telecommunications engineering in 1993 and the PhD degree in the same field in 2001 from Universidad Politécnica de Madrid, Spain. He is a professor (since 2003) at the Universidad Autonóma de Madrid, where he codirects the Video Processing and Understanding Lab. His research interests include the analysis of video sequences, video indexing based on content, 2-D and 3-D machine vision.