International Journal of Bioelectromagnetism *www.ijbem.org*
Vol. 14, No. 4, pp. 217 - 222, 2012

Extending the Universal Quality Index to Assess N-image Fusion in Light Microscopy

Filippo Piccinini[a], Anna Tesei[b], Wainer Zoli[b], Alessandro Bevilacqua[ac]
[a]*ARCES, Advanced Research Center on Electronic Systems, University of Bologna, Italy*
[b]*IRCCS-IRST, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori, Meldola (FC), Italy*
[c]*DISI, Department of Computer Science and Engineering, University of Bologna, Italy*

Correspondence: A Bevilacqua, Advanced Research Center on Electronic Systems, University of Bologna, Via Toffano 2/2, IT-40125 Bologna, Italy. E-mail: alessandro.bevilacqua@unibo.it, phone +39 051 2095409, fax +39 051 2095410

Abstract. The *depth of focus* is a constraint of all the light microscopes, being not possible to acquire a single, completely in-focus, image of 3D objects characterized by a wide depth dimension. Since the seventies, many methods have been proposed to extend the microscope's depth of focus. However, it is difficult to define which method yields the best result because there is not a tested metric and a validated approach to compare them using a set of images lacking of a reference ground truth. Typically, the Universal Quality Index (UQI) is used to evaluate output's quality in image processing, but it requires a reference ground truth. Some UQI extensions were proposed to evaluate the output of fusion methods without a ground truth, but no exhaustive analysis was proposed to prove their equivalence, in terms of evaluation performance, to the standard UQI. In this work, first we prove that the output of the proposed UQI extensions is different from the UQI's one, then we provide and validate a new approach to evaluate extended depth of focus methods, such as the UQI would do, but without requiring a reference ground truth.

Keywords: Image Processing, Depth of Field, Microscopy Image Fusion, No-Reference Metric, Universal Quality Index

1. Introduction

The objects observed to the light microscope are often thicker than the *depth of focus* (also known as *depth of field*), meaning that it is not possible to acquire them in a single completely focused image. This is a well known problem in biological imaging performed with all the widefield microscopes. For instance, in case of acquisition of a large spheroidal agglomerate of cells, it would be impossible not to have out-focus regions [Tesei et al., 2009].

Many methods have been proposed in literature to extend the microscopes' depth of field through digital processing [Aguet et al., 2008; Forster et al., 2004; Goldsmith, 2012; Piccinini, Tesei et al., 2012; Tessens et al., 2007; Valdecasas et al., 2001]. Usually, all these methods start by exploiting a *stack* of images acquired by moving the microscope's holder along the principal axis (z-direction), in order to obtain optical sections (*slices*) of the specimen. They rely on the assumptions that the slices are perfectly z-aligned and the acquired objects are not in motion. In particular, the extended depth of focus algorithms aim to recover the in-focus regions from each single slice in order to build a final single composite image, more informative and suitable for visual inspection or quantitative analyses. Furthermore, the fusion process could be an introductory stage to innovative tasks, like 3D reconstruction and densitometric measurements, where the performance of the fusion algorithm strongly affects the accuracy of the entire process. A survey of literature image fusion methods is provided in [Valdecasas et al., 2001].

Due to the huge number of extended depth of focus methods proposed in literature and the lack of validated approaches to compare them, it is hard to determine which ones could best fit one's application. *Universal Quality Index* (UQI), proposed in [Wang and Bovik, 2002], is a wide employed metric to numerically compare images. Although UQI is not explicitly based on any human visual system, the authors in [Wang and Bovik, 2002] proved that it represents the image's quality more closely to the human visual perception. This metric requires a reference *ground truth* to be evaluated and in real world image fusion applications this is often a missing information. To overcome this problem, a widely used extension of the UQI metric is proposed in [Piella and Heijmans, 2003]. This metric, known as *Piella's Metric* (PM), does not require any reference and it permits to compare methods based on fusion of *only two* information sources. Finally, a further extension of PM is proposed in [Bueno et al., 2005]. This metric (hereinafter, referred to as UQIN) is suitable to compare performances of methods where *N* images are used to obtain the final composite one. Both PM and UQIN metrics, claimed to be UQI's extensions, are assumed to be able to numerically compare the output such as the original UQI would do. However, no exhaustive analysis was performed to confirm this correlation.

In this work, through an extensive analysis we prove that UQIN provides very different rankings from those achieved using the original UQI. To this aim, we employed five different extended depth of focus methods, using stacks of synthetic slices built on purpose, where the original images exploited to obtain the stacks are used as the ground truth as well. The two different rankings obtained prove that the UQIN and the standard UQI are not equivalent in terms of quality evaluation. Finally, we provide a new approach to numerically evaluate fusion methods of *N* images, without a reference ground truth, capable to perform as the original UQI.

2. Material and Methods

Many methods have been proposed in literature to obtain a composite in-focus image starting from a stack of partially-unfocused ones. In this work, five different extended depth of focus tools are compared. Four of them are widely employed and their implementation freely distributed inside ImageJ (http://rsbweb.nih.gov/ij/), one of the most common open source program used for image processing [Bulj et al., 2013]. *Stack Focuser* (SF) and *Depth From Focus* (DFF) can be downloaded from the ImageJ Plug-In web section. Both work in the spatial domain and are based on different decision rules to determine the in-focus regions on the images of the stack: substantially, SF is based on an edge detector stage, while DFF uses a contrast metric, named sharpness index [Goldsmith, 2012]. The remaining two ImageJ software used in the comparison tests are available from the website http://bigwww.epfl.ch/demo/edf/ of the Biomedical Imaging Group (BIG) of the École Polytechnique Fédérale de Lausanne (EPFL). The first one implements the *Complex Wavelet-Based Method* (CWBM) described in [Forster et al., 2004] and exploits the complex wavelet transform to locally analyze and define the in-focus regions of each single image of the stack. The second one, implementing the *Model-Based Method* (MBM) explained in [Aguet et al., 2008], faces the extension of the depth of field as a deconvolution optimization problem, where object's texture and topography are jointly estimated in an iterative process. The last tested software works in the frequency domain exploiting the *curvelet* transform to separate the high frequency image content to reconstruct the final image from the in-focus regions [Tessens et al., 2007]. Hereinafter we refer to the method as CUR. CUR has been implemented in MATLAB© (The MathWorks, Inc., Massachusetts, USA) using functions available from the website www.curvelet.org and our implementation can be provided on demand.

To evaluate whether the UQIN metric is able to classify the methods like the standard UQI, five synthetic stacks (of eight images each) were constructed starting from five images, with very different contents, of the Brodatz textures database (http://www.ux.uis.no/~tranden/brodatz.html). In particular, we used images D13, D22, D23, D61, D112 (Fig. 1). D13 and D61 can be considered similar to bone tissue, D22 to a lung biopsy, and D61 and D112 to cell cultures. Each stack of slices was created by projecting the Brodatz texture on a 3D surface and convolving it with a 3D point spread function. Finally, the corresponding volume was linearly interpolated along the *z*-direction to obtain a *z*-stack of synthetic slices. In particular, in order to build the synthetic stacks of partially-unfocused slices simulating the acquisition at different focal planes, we exploited the specific function available in the MBM interface and already used in [Aguet et al., 2008].

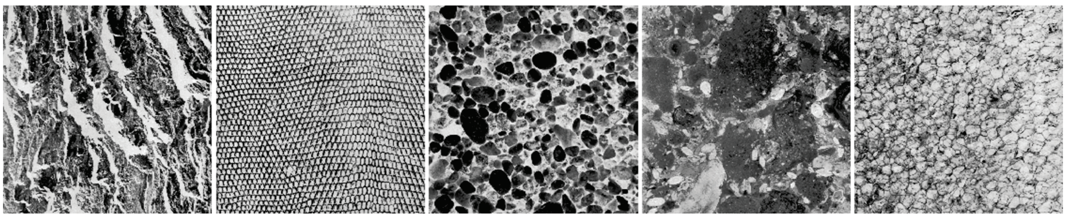

Figure 1. *Gray images from the Brodatz textures database. From left to right: D13, D22, D23, D61, D112.*

In order to obtain a performance ranking of the methods under testing and compare them, the output in-focus images reconstructed from the synthetic stacks were evaluated using both UQI and UQIN. The standard UQI, defined accordingly to Eq. 1:

$$UQI(g,f)=\left(\frac{\sigma_{gf}}{\sigma_g\cdot\sigma_f}\right)\left(\frac{2\cdot\overline{x_g}\cdot\overline{x_f}}{\overline{x_g}^2+\overline{x_f}^2}\right)\left(\frac{2\cdot\sigma_g\cdot\sigma_f}{\sigma_g^{\ 2}+\sigma_f^{\ 2}}\right) \tag{1}$$

is designed by modelling any image distortion as a combination of three factors: loss of correlation, luminance and contrast distortion [Wang and Bovik, 2002].

$\overline{x_g}$, $\overline{x_f}$, σ_g , σ_f and σ_{gf} are mean, standard deviation and covariance, respectively, of the images *g* and *f*, where *g(x,y)* is the ground truth and *f(x,y)* is the final composite image obtained from the stack. The basic version of the UQIN metric is reported in Eq. 2:

$$UQIN(s_1,\ldots,s_N,f)=\frac{1}{|T|}\sum_{w\in T}\sum_{n=1}^{N}\lambda_{s_n}(w)\cdot UQI(s_n,f\mid w) \tag{2}$$

s_n is one of the *N* slices used in the metric evaluation (for PM, *N*=2), *w* is a small ROI region of the collection *T* and $\lambda_{s_n}(w)$ is defined accordingly to Eq. 3:

$$\lambda_{s_n}(w)=\frac{Y(s_n\mid w)}{\sum_{n=1}^{N}Y(s_n\mid w)} \tag{3}$$

where $Y(s_n\mid w)$ is the local saliency of the image s_n , typically defined as the image's variance in the ROI *w*.

3. Results

To compare the extended depth of focus methods, UQI and UQIN are evaluated using the in-focus images obtained from the five synthetic stacks. The original Brodatz's images are used as the reference ground truth to assess UQI, while all the images of each stack are used in place of the ground truth for UQIN. For each stack, Tab. 1 reports the software ranking, computed by assigning the 1st position to the method that obtained the best result (i.e., having the highest value for both UQI and UQIN). It is worth noting that the ranking obtained by the standard UQI is the same for all the stacks used, while using UQIN the ranking is image-dependent. Despite UQIN is proposed as an extension of the standard UQI, our results prove that the two metrics do not provide the same ranking. The only constant position is the 5th one, always achieved by DFF. The worst position is also confirmed by a visual analysis: while it is very difficult by looking at the five in-focus images to say which software yields the best result, the worst result can be easily assessed (Fig. 2).

Table 1. *Ranking based on UQI and UQIN.*

	MBM	*CWBM*	*DFF*	*SF*	*CUR*
UQI all stacks	4^{th}	2^{nd}	5^{th}	1^{st}	3^{rd}
UQIN D13	1^{st}	3^{rd}	5^{th}	4^{th}	2^{nd}
UQIN D22	1^{st}	4^{th}	5^{th}	3^{rd}	2^{nd}
UQIN D23	1^{st}	3^{rd}	5^{th}	4^{th}	2^{nd}
UQIN D61	1^{st}	4^{th}	5^{th}	3^{rd}	2^{nd}
UQIN D112	1^{st}	4^{th}	5^{th}	3^{rd}	2^{nd}

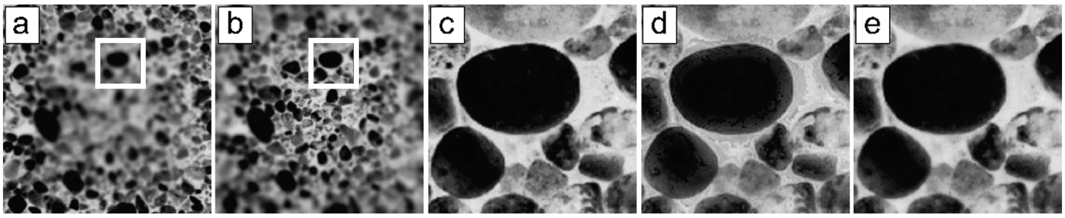

Figure 2. ***a*** *and* ***b****: partially-unfocused slices of the stack synthetically built using the image D23.* ***c****: a zoomed detail of the ground truth image D23, related to the ROI surrounded by the white square highlighted in* ***a*** *and* ***b****.* ***d*** *and* ***e****: same detail of* ***c****, but from the images obtained using DFF and ST, respectively.*

For the sake of completeness, besides the basic UQIN version reported in Eq. 2, we also tested the versions described in [Bueno et al., 2005], but without ever achieving the same ranking of UQI.

To be certain that the standard UQI could be used to obtain a fair ranking method we also computed the Mean Squared Error (MSE) and the Signal to Noise Ratio (SNR) between the obtained in-focus images and the ground truth (entities typically used in image processing – formulas and further details can be found in [Piccinini, Lucarelli et al., 2012]). These two quality metrics are the most used in image processing and often the rankings achieved by these metrics and UQI are the same [Bevilacqua et al., 2011; Carozza et al., 2011; Gherardi et al., 2011; Piccinini, Lucarelli et al., 2012]. Tab. 2 reports the MSE and the SNR values obtained. As it can be seen, for all the images the ranking is the same achieved using the standard UQI. It is worth noting that the values yielded by DFF are by far the worst (highest MSE and lowest SNR), while the values achieved by the other methods are quite similar. This confirms that, through visual analysis only, just the method leading to the worst result can be easily detected.

From a visual point of view, the worst final composite images may contain evident noise, Airy discs, or artefacts (e.g., produced by diffraction and aliasing [Goldsmith, 2012], Fig. 2d). Starting from the consideration that it is quite easy to visually decree whether a method yields bad results, we propose a new approach to evaluate the methods without exploiting any reference ground truth. By its definition, the standard UQI aims at quantifying *how similar* two images are. Ideally, an in-focus image obtained from an extended depth of focus method should be *corrected* for the blurring effect within the defocused regions in each slice of the processed stack. Accordingly, the more effective is the extended depth of focus method, the higher would be the contribution, in terms of blur compensation, retained by the final composite image with respect to each slice. By quantifying all these contributions in terms of *image dissimilarity*, a better-quality final composite image should be *distant*, on the average, from the processed slices, thus yielding a higher image dissimilarity score, accordingly. Following these assumptions, we propose an effective approach to estimate the ranking of extended depth of focus methods, with the same evaluation performance of the standard UQI, which does not require a reference ground truth. First, based on the visual quality of the final images, we discard the methods clearly yielding poor results. Second, for the remaining methods, the Average UQI (hereinafter, AUQI) between the obtained composite in-focus image and the slices of the original stack is computed according to Eq. 4:

$$AUQI = \frac{1}{N}\sum_{n=1}^{N} UQI(s_n, f) \tag{4}$$

Finally, the ranking of methods is obtained by attributing the lowest AUQI value to the best method, and so on. In this way, the worst method is the one obtaining higher AUQI value, this meaning that the final in-focus image is more similar to the original blurred slices. It is worth noting that the AUQI, like the standard UQI, is proposed as a global index without any weighting function, while the UQIN is defined as a weighted local index and it could play an important role in the image ranking.

In Tab. 3, for each stack the AUQI values and the final ranking obtained are reported. According to the proposed approach, we discarded the method DFF, that clearly always gives the worst results. In particular, the ranking obtained is the same for all the stacks and it is equivalent to the one achieved using the standard UQI.

Table 2. *MSE, SNR and ranking.*

	MBM	*CWBM*	*DFF*	*SF*	*CUR*
MSE D13	*77.97*	*57.17*	*291.44*	*53.38*	*68.69*
MSE D22	*143.50*	*100.23*	*415.05*	*97.53*	*137.62*
MSE D23	*32.58*	*24.85*	*244.39*	*23.43*	*29.88*
MSE D61	*23.14*	*18.97*	*153.26*	*16.94*	*21.45*
MSE D112	*54.29*	*42.35*	*281.93*	*39.02*	*51.29*
SNR D13	*24.68*	*26.03*	*18.95*	*26.32*	*25.23*
SNR D22	*22.71*	*24.27*	*18.10*	*24.39*	*22.89*
SNR D23	*27.46*	*28.63*	*18.71*	*28.89*	*27.83*
SNR D61	*28.11*	*28.98*	*19.90*	*29.47*	*28.44*
SNR D112	*27.50*	*28.58*	*20.35*	*28.94*	*27.75*
Ranking	*4^{th}*	*2^{nd}*	*5^{th}*	*1^{st}*	*3^{rd}*

Table 3. *AUQI and ranking.*

	MBM	*CWBM*	*DFF*	*SF*	*CUR*
AUQI D13	*0.8917*	*0.8831*	*//*	*0.8823*	*0.8896*
AUQI D22	*0.6688*	*0.6542*	*//*	*0.6540*	*0.6673*
AUQI D23	*0.9273*	*0.9229*	*//*	*0.9222*	*0.9269*
AUQI D61	*0.9359*	*0.9314*	*//*	*0.9308*	*0.9354*
AUQI D112	*0.8363*	*0.8257*	*//*	*0.8252*	*0.8352*
Ranking	*4^{th}*	*2^{nd}*	*5^{th}*	*1^{st}*	*3^{rd}*

4. Conclusions

In this work we faced the problem of evaluating extended depth of focus methods in cases of no reference ground truth is at one's disposal. The tests performed using synthetic stacks proved that the metrics proposed in literature as UQI extensions are not able to rank the methods as the standard UQI does. We propose a new approach to evaluate extended depth of focus methods without requiring a ground truth and exploiting all the *N* images of the stack. It works according to the following steps. First, the worst methods are discarded through visually evaluating the obtained in-focus images. Second, the average UQI is calculated between the final in-focus images and all the slices of the stack. Finally, the ranking is performed by attributing the 1^{st} position to the method yielding the most different final image with respect to the original blurred slices, and accordingly the last position to the most similar final image. The experiments proved that the rankings obtained following our new approach are the same achieved by the standard UQI. This confirms that our approach can be used to evaluate extended depth of field methods in real world cases, where N images are recorded and no ground truth is at one's disposal.

The implementations of the methods and all the metrics used in the tests can be provided on demand.

Acknowledgements

We thank Alessandro Gherardi (Computer Vision Group, University of Bologna, Italy) and Enrico Lucarelli, jointly with all his staff of the Osteoarticular Research Laboratory (Istituto Ortopedico Rizzoli, Bologna, Italy), for helpful discussions. Ludovico Carozza (School of the Built Environment at Herriot-Watt University, Edinburgh, United Kingdom) and Davide Pollini (Master Graduate Student, Biomedical Engineering School of Cesena, University of Bologna, Italy) for their contribution provided in the initial part of this work.

References

Aguet F, Van de Ville D, Unser M. Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy. *IEEE Transactions on Image Processing*, 17(7): 1144-1153, 2008.

Bevilacqua A, Piccinini F, Gherardi A. Vignetting correction by exploiting an optical microscopy image sequence. In Proceedings of the 33th Annual International Conference of the IEEE/EMBS, 2011, 6166-6169.

Bueno MA, Álvarez-Borrego J, Acho L, Chávez-Sánchez MC. Polychromatic image fusion algorithm and fusion metric for automatized microscopes. *Optical Engineering*, 44(9): 093201.1-093201.7, 2005.

Bulj Z, Duchi S, Bevilacqua A, Gherardi A, Dozza B, Piccinini F, Mariani GA, Lucarelli E, Giannini S, Donati D, Marmiroli S. Protein kinase B/AKT isoform 2 drives migration of human mesenchymal stem cells. *International Journal of Oncology*, 42(1): 118-126, 2013.

Carozza L, Bevilacqua A, Piccinini F. Mosaicing of optical microscope imagery based on visual information. In Proceedings of the 33th Annual International Conference of the IEEE/EMBS, 2011, 6162-6165.

Forster B, Ville DVD, Berent J, Sage D, Unser M. Complex wavelets for extended depth-of-field: a new method for the fusion of multichannel microscopy images. *Microscopy Research and Technique*, 65(1-2): 33-42, 2004.

Gherardi A, Bevilacqua A, Piccinini F. Illumination field estimation through background detection in optical microscopy. In Proceedings of the 8th annual IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2011, 49-54.

Goldsmith NT. Deep focus; a digital image processing technique to produce improved focal depth in light microscopy. Image Analysis & Stereology, 19(3): 163-167, 2011.

Piccinini F, Lucarelli E, Gherardi A, Bevilacqua A. Multi-image based method to correct vignetting effect in light microscopy images. *Journal of Microscopy*, 248(1): 6-22, 2012.

Piccinini F, Tesei A, Zoli W, Bevilacqua A. Extended depth of focus in optical microscopy: assessment of existing methods and a new proposal. *Microscopy Research and Technique*, 75(11): 1582-1592, 2012.

Piella G, Heijmans H. A new quality metric for image fusion. In Proceedings of the 10th International Conference on Image Processing, 2003, 173-176.

Tesei A, Zoli W, Arienti C, Storci G, Granato AM, Pasquinelli G, Valente S, Orrico C, Rosetti M, Vannini I, Dubini A, Dell'Amore D, Amadori D, Bonafè M. Isolation of stem/progenitor cells from normal lung tissue of adult humans. *Cell Proliferation*, 42(3): 298-308, 2009.

Tessens L, Ledda A, Pizurica A, Philips W. Extending the depth of field in microscopy through curvelet-based frequency-adaptive image fusion. In Proceedings of the 32th IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, 861-864.

Valdecasas A, Marshall D, Becerra J, Terrero J. On the extended depth of focus algorithms for bright field microscopy. *Micron*, 32(6): 559-569, 2001.

Wang Z, Bovik AC. A universal image quality index. *IEEE Signal Processing Letters*, 9(3): 81-84, 2002.