Università degli studi di Udine

Generation of All-in-Focus Images by Noise-Robust Selective Fusion of Limited Depth-of-Field Images

Original

Availability:
This version is available http://hdl.handle.net/11390/890946 since 2016-01-13T15:25:11Z

Publisher:

Published
DOI:10.1109/TIP.2012.2231087

Terms of use:
The institutional repository of the University of Udine (http://air.uniud.it) is provided by ARIC services. The aim is to enable open access to all the world.

Publisher copyright

(Article begins on next page)

12 October 2024

Generation of All-in-focus Images by Noise-robust Selective Fusion of Limited Depth-of-field Images

Said Pertuz, Domenec Puig, Miguel Angel Garcia, Andrea Fusiello

***Abstract*—The limited depth-of-field of some cameras prevents them from capturing perfectly focused images when the imaged scene covers a large distance range. In order to compensate for this problem, image fusion has been exploited for combining images captured with different camera settings, thus yielding a higher quality *all-in-focus* image. Since most current approaches for image fusion rely on maximizing the spatial frequency of the composed image, the fusion process is sensitive to noise. In this work, a new algorithm for computing the all-in-focus image from a sequence of images captured with a low depth-of-field camera is presented. The proposed approach adaptively fuses the different frames of the focus sequence in order to reduce noise while preserving image features. The algorithm consists of three stages: focus measure, selectivity measure and image fusion. An extensive set of experimental tests has been carried out in order to compare the proposed algorithm with state-of-the-art all-in-focus methods using both synthetic and real sequences. The obtained results show the advantages of the proposed scheme even for high levels of noise.**

***Index Terms*—All-in-focus, image fusion, extended depth of field, focus measure**

I. Introduction

THE limited depth of field of optical systems is a common problem in image acquisition since it leads to the defocusing of those parts of the depicted scene that are not comprised within the in-focus limits, hence affecting the quality and amount of information that can be retrieved from the captured images. In particular, when images are captured with a low depth-of-field system, the objects at different depths from the camera may appear out of focus if they are away from the focus plane.

Image fusion aims at combining several images of the same scene with different camera settings and a same distance between the camera and the scene. The aim is to obtain a higher quality image more suitable for human or machine interpretation (for instance, focus fusion, high dynamic range creation, etc). In the particular case of low depth-of-field imaging, image fusion can be carried out by combining images of a focus sequence, that is, a sequence of images of the same scene captured with different focus settings. The goal is to obtain an image where all the visible objects are focused. Such an image is referred to as an all-in-focus (AIF) image. This paper assumes that the different frames of the focus sequence are properly aligned, that is, that the image shift due to parallax or magnification change is negligible or has already been compensated.

Previous algorithms proposed in the literature to compute the AIF image can be broadly organized into four main families: methods based on the spatial frequency, image pyramids, defocus modeling and wavelet transforms. Methods based on the *spatial frequency* usually apply a sharpness measure or focus measure in order to identify the pixels with higher information content in each frame [1]. The methods based on *image pyramids* usually perform a multi-scale decomposition of the image in order to identify the pixels or image regions with higher information content at different scales [2], [3]. Alternatively, the methods based on *defocus modeling* recover the AIF image under the assumption of a known point spread function model and then apply a filter designed to reverse its effect [4]–[6]. In order to work appropriately, these methods rely on an estimation of the parameters of the point spread function. Finally, the methods based on the *wavelet transform* carry out a wavelet decomposition of the focus sequence. The image fusion is then performed in the wavelet domain by selecting the wavelet coefficients according to some criterion [1], [7]. The wavelet transform can also be considered to be as an instance of a multi-scale decomposition, resembling the approaches based on image pyramids, although the coefficients are selected in the wavelet domain instead of in the spatial domain.

In general, most all-in-focus methods, with the exception of defocus modeling-based methods, can be described through the following *energy maximization scheme* [6]:

1) An image stack, $I_k(x, y)$, is acquired, where (x, y) denotes the spatial coordinates and k is the frame number. Each I_k corresponds to an image captured at a certain in-focus distance.
2) Either a high frequency measure or a focus measure is applied to each frame of the image stack in the space domain, the scale-space domain or the wavelet domain depending on the method applied in order to compute the AIF image.
3) An index map is generated such that each position (x, y) keeps the index z of the frame with the largest frequency or focus measure for that position.
4) The all-in-focus image is generated based on the previous index map. In the particular case of pyramid-based

Manuscript received October 11, 2011; revised May 11, 2012; accepted October 29, 2012.

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

S. Pertuz and D. Puig are with the Dept. of Computer Science and Mathematics, Universitat Rovira i Virgili, Tarragona, Spain. e-mail: {said.pertuz,domenec.puig}@urv.cat

M. A. Garcia is with the Dept. of Informatics Engineering, Autonomous University of Madrid, Madrid, Spain. email: miguelangel.garcia@uam.es

A. Fusiello is with the Dept. of informatics, Università degli Studi di Verona, Verona, Italia.

or wavelet-based methods, an inverse transformation is usually required.

An important drawback of the scheme described above is that, in the presence of noise, the maximization of the focus measure will also add noise to the final result. The effect of noise can be attenuated by applying a low-pass filter in either a pre- or post-processing step at the expense of image contrast.

In this paper, an alternative robust algorithm for computing the AIF image in the presence of noise is presented. The algorithm consists of three stages: focus measure, selectivity measure and image fusion. This algorithm takes advantage of the effect of noise in the third dimension of the image stack in order to perform a selective fusion of images. An extensive set of experimental tests has been carried out in order to identify its advantages and weaknesses. In addition, a comparative study with state-of-art algorithms shows the benefits of the proposed approach.

This paper is organized as follows: the next section describes the proposed selective all-in-focus algorithm (SAF). In section III, the proposed method is compared with state-of-the-art algorithms using synthetic and real focus sequences. Finally, a discussion of the obtained results and conclusions are presented in sections IV and V, respectively.

II. Proposed approach

According to the theory of defocus, the irradiance of a point spreads as it departs from the in-focus position [8]. Therefore, focus measure algorithms estimate the local energy of the processed images. This principle is valid for the detection of the in-focus position in autofocus [4], shape recovery in shape-from-focus [9] and all-in-focus computation in image fusion.

The maximization of energy for detecting maximum focus remains valid in the ideal case of noiseless images. Notwithstanding, the relationship between signal strength and noise level in real images is critical in order that changes in focus be detectable by traditional methods.

In this section, a new methodology to recover the all-in-focus image from a focus sequence is presented. The proposed selective image fusion algorithm adapts the fusion rule to the strength of the signal with respect to noise. This allows for a selective noise filtering in the spatial domain that preserves image features in highly textured areas. The selective fusion is performed in three main steps: focus measure, selectivity measure and image fusion.

A. Focus Measure

An important step in image fusion is the application of an activity measure or focus measure. Focus measurement has been an intensive research field for 3D shape recovery in *shape-from-focus* and *autofocus*. Focus measure operators often work by applying a transformation to the original image in order to enhance its sharpness. The energy of the transformed image over a region of interest is then used as a focus level estimator. Many focus measure algorithms that operate in both the spatial and frequency domains have been proposed. In the case of spatial domain operators, images are usually divided into blocks of fixed size (e.g., [10]). The aim of a block is to capture enough image information in order to detect changes in focus. The block size should be large enough to allow an accurate measurement of the focus level and small enough to only encompass regions with a similar focus degree [11].

A different approach to block-wise computation is the application of pixel-wise measures (e.g., [12]). In this case, a focus measure operator is computed for every pixel by taking into account a small neighborhood around that pixel as a support region. A focus measure matrix, F_k, is computed for each image I_z of the stack. The values of the focus measure for a pixel at coordinates (x, y) over all the image frames are referred to as *focus function* (or focus measure vector): $f_{x,y} = (F_1(x, y), ...F_k(x, y), ...F_N(x, y))$, where $F_k(x, y)$ is the focus measure of that pixel at the k-th frame and N is the total number of frames.

In the literature, several focus measure operators have been proposed based on different working principles, such as the image gradient (gradient energy [9] and Tenengrad algorithm [13]), image Laplacian (the modified Laplacian [14] and energy of Laplacian [4]), and image statistics (the gray-level variance [15]), among others.

A comparison of different focus measure operators for autofocus can be found in [13]. Any of the aforementioned focus measure operators could be applied to the present application. In particular, a pixel-wise measure, the gray-level variance has been used as a focus measure in this work:

$$F_k(x, y) = \sum_{(i,j)\in\Omega(x,y)} (I(i, j) - \mu)^2, \tag{1}$$

where $\Omega(x, y)$ is the $r \times r$ neighborhood of (x, y) and μ is the mean gray-level of pixels within $\omega(x, y)$. The selection of r is a trade-off between robustness to noise and spatial resolution [16] and is, hence, application-dependent. A radius $r = 9$ has experimentally been defined in this work.

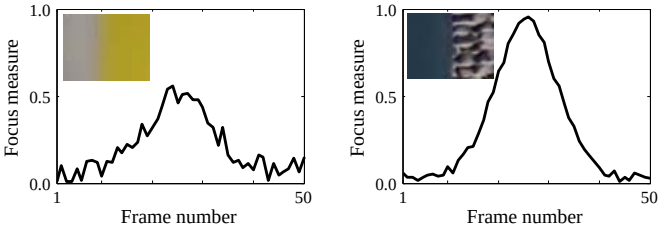
B. Selective measure

As all focus measures, the measure in (1) is also sensitive to noise. In order to perform a selective image fusion for generating a low-noise all-in-focus image, the aforementioned focus measure is complemented with a selection scheme that allows the system to determine if the focus measurement is reliable or not. The proposed approach is illustrated in Fig. 1, where the focus functions corresponding to different image regions are plotted. From those curves, it can be appreciated that the focus functions corresponding to regions with different texture patterns exhibit different behaviors. On the one hand, the image features in Fig. 1(a) are "weak" when compared to the noise present in the sequence. On the other hand, the image texture in Fig. 1(b) stands out over the existing noise, leading to a clear response with a maximum at the position of highest focus.

Let the focus function for a pixel at coordinates (i, j) be a signal that varies according to both the degree of focus of this pixel and some additive noise:

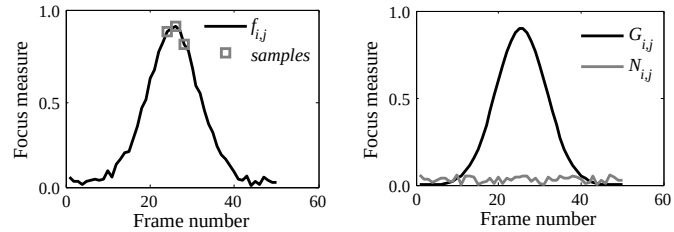
$$f_{i,j} = G_{i,j} + N_{i,j}, \tag{2}$$

where $f_{i,j}$ is the computed focus function for that pixel, $G_{i,j}$ the associated ideal focus function and $N_{i,j}$ a noise signal



(a) (b)

Fig. 1. Behavior of focus measures for different texture patterns. (a) Low-textured patterns. (b) Highly-textured patterns



(a) (b)

Fig. 2. Decomposition of the focus measure function. (a) Function sampling for Gaussian fit. (b) Decomposition into $G_{i,j}$ (signal) and $N_{i,j}$ (noise) components.

that represents the departure of the focus function from the ideal behavior. $N_{i,j}$ may be explained by image noise, lack of texture and limitations of the focus measure operator. By studying the behavior of the focus function, it is reasonable to model $G_{i,j}$ as a Gaussian function. This assumption is often applied in shape-from-focus for the computation of topology maps or depth-maps from focus sequences (e.g., [14]).

Let the ideal Gaussian function corresponding to the pixel at coordinates (i, j) be defined as:

$$G_{i,j}(z) = A \exp\left(-(z-\mu)^2/(2\sigma^2)\right), \tag{3}$$

where A is the maximum value of the Gaussian function, μ its mean value and σ its standard deviation. Following [14], A, μ and σ are found by interpolation as a function of the depth value z. As shown in Fig. 2(a), three points around the maximum value of $f_{i,j}$ are used in order to fit the above Gaussian. Once $G_{i,j}$ has been estimated, the noise signal $N_{i,j}$ is readily computed from (2). The selectivity measure for the pixel at coordinates (i, j), namely $S(i, j)$, is obtained from $G_{i,j}$ and $N_{i,j}$ as the PSNR:

$$S(i,j) = 20 \log(\max(f_{i,j})/\text{RMS}(N_{i,j})). \tag{4}$$

This selectivity measure is intended to describe how well the focus measure adjusts to the ideal behavior, and determines if the image fusion at the next step should give priority either to the preservation of image features or to the reduction of the effects of noise. The PSNR of $f_{i,j}$ will yield low values for pixels corresponding to regions with weak features where the focus measure operator fails to measure focus accurately and, thus, where the effects of noise are not negligible.

The Gaussian fit illustrated in Fig. 2 corresponds to an idealized *focus profile*. The focus profile refers to the shape of the focus function and is an important concept studied in autofocus and shape-from-focus. On the one hand, it depends on the intensity distribution of the image, as a function of the amount of defocus and the response of the focus measure operator to those intensities. Therefore, the real focus profile depends on both the scene luminance and the focus operator. On the other hand, it depends on the characteristics of the imaging device (in terms of the focal length, numerical aperture and resolution) and the scene geometry. Thus, without prior knowledge about these variables, it is difficult to derive the exact shape of the focus profile. In addition to the Gaussian fit [14], some researchers have also proposed quadratic or polynomial fits [4]. Recently, Mannan and Choi [17] proposed a new focus profile model by only considering the intensity distribution of the focused image along the optical axis; and Tsai and Chen [18] derived a focus profile model for a particular focus measure operator based on the horizontal gradient of the image.

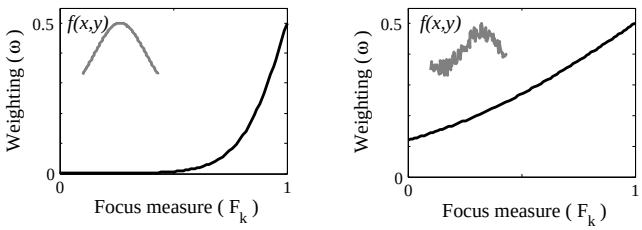
Gaussian interpolation has been used in this work for its generality (without assumptions on the applied focus measure operator, the luminance of the scene or the configuration of the acquisition device) and since it correctly describes well the desired behavior of the focus profile: a sharp and noiseless peak. Whenever the focus profile departs from this ideal behavior, either due to the presence of noise or the acquisition conditions, the selective measure in (4) will respond accordingly. The suitability of the proposed fit for this particular application is experimentally assessed in section III.

C. Image fusion

A straightforward solution to the image fusion problem is to compute the intensity of a pixel in the all-in-focus image ψ at coordinates (x, y) as a weighted average of the pixels in the original sequence. The weights will be proportional to the activity of the image pixels (the focus measure value): $\psi(x, y) = \sum_{k=1}^{K} \tilde{F}_k(x, y) I_k(x, y)$, where $\tilde{F}_k$ is the focus measure normalized so that $\sum_k \tilde{F}_k = 1$. Notwithstanding, this approach has two drawbacks: first, a linear combination of all frames yields a low-contrast, all-in-focus image. Second, the sensitivity of all focus measure operators to high-frequency components in the images will yield a low-quality all-in-focus image in the presence of noise. Bearing this in mind, in the final step of the proposed SAF algorithm, an image fusion process is performed according to the activity of image pixels, the later estimated by means of both the focus measure (1) and the relevance of the image features estimated through the selectivity measure (4):

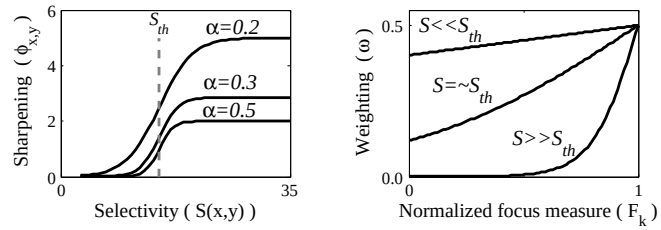
$$\psi = \frac{1}{\Omega} \sum_{k=1}^{K} \omega(k) I_k, \tag{5}$$

where $\Omega = \sum_{k=1}^{K} \omega_k$ and $\omega(k)$ are weighting coefficients adaptively computed based on the image content by applying a transfer function to the focus measure. This transfer function is modulated by the selectivity measure. Thus, the overall transformation will adapt to the image content. The definition



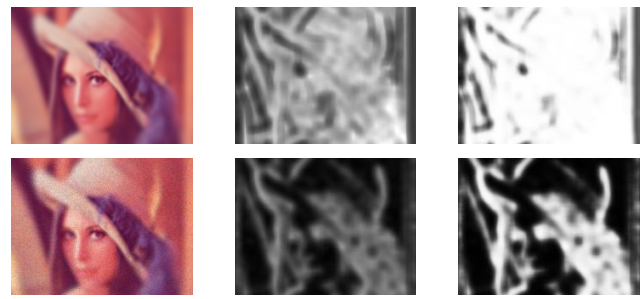
(a) (b)

Fig. 3. Selective weighting. (a) Weighting for an ideal focus function. (b) Weighting for a noisy focus function.



(a) (b)

Fig. 4. (a) Effect of α on the strength of selectivity. (b) Transfer function for different values of S.



(a) (b) (c)

Fig. 5. Working principle of the proposed SAF algorithm. (a) Defocused frame. (b) Selectivity measure $S(x, y)$. (c) Sharpening parameter $\phi_{x,y}$

of those coefficients is fully described below. For simplicity, sub-indices (x, y) have been omitted in (5). In the sequel, all symbols and equations refer to pixels at coordinates (x, y) unless otherwise is indicated.

The intensities in the AIF image corresponding to pixels that exhibit a strong visual pattern (nearly ideal behavior) are generated by giving a larger weight to the intensities of those pixels with a higher focus measure. In other words, the energy maximization scheme is held for pixels with a high selectivity measure. In contrast, the intensities of the pixels that exhibit weak visual patterns and hence have a greater noise influence are generated by averaging the original intensities over the whole focus sequence, thus giving preference to noise reduction. The objective of the transfer function is to provide a smooth continuous transition between these two extreme cases.

For illustration purposes, Fig.3 shows the desired weights in (5) as a function of the normalized focus measure value for two different cases. Fig. 3(a) shows the weights for an idealized focus measure function $f(x, y)$ that gives preference to those pixels with the highest focus values ($F_k \to 1$). In contrast, Fig. 3(b) shows the weights for a non-ideal focus function. In this case, pixels with lower focus value still have a significant contribution on the computation of the AIF image.

The behavior of ω in Fig. 3 is analogous to a high-pass filter in the frequency domain. In this work, this analogy has been exploited in order to propose a definition of ω. Digital filters are designed to cope with certain desired characteristics in the frequency domain while keeping an efficient time domain representation. Alternatively to traditional FIR and IIR filters, the hyperbolic tangent-based filters allow easy control of the cut-off frequency and the transition band [19], [20]. A general pass-band hyperbolic tangent-based filter is defined in the frequency domain of f as :

$$H(f) = \frac{\tanh(\phi(f \pm f_c)) + 1}{2}, \tag{6}$$

where f_c determines the band-pass frequency and ϕ controls the transition band (slope).

Equation (6) is suitable for the sought weighting in (5) since it has a fast exponential decay (e^{-f}) for points away from the cut-off frequencies and can be easily parameterized as a function of ϕ. In particular, since the normalized focus measure is between 0 and 1, ω is defined as a high-pass filter with a cut-off frequency of 1 [20]:

$$\omega(k) = \frac{1}{2} + \frac{1}{2}\tanh(\phi(F_k - 1)). \tag{7}$$

Equation (7) is a logistic function that provides a continuous transition between the minimum and maximum values of F_k. The speed of that transition is modulated by the *sharpening parameter* ϕ which, in turn, is a sigmoid function of the selectivity:

$$\phi = \frac{1}{2\alpha}(1 + \tanh(\alpha(S - S_{th}))). \tag{8}$$

The transformation described in (7) and (8) depends on two parameters: a selectivity threshold (in decibels), $S_{th} \in [0, \infty]$, and a selectivity constant, $\alpha \in (0, 1]$. These parameters can be described as follows (see Fig. 4): the selectivity threshold indicates the inflexion point at which preference is given either to the preservation of image features ($S \gg S_{th}$) or to the suppression of artifacts ($S \ll S_{th}$). The selectivity constant allows the user to specify the *strength* of the selectivity in the fusion process. A value of $\alpha \to 0$ will lead to a higher selectivity. The sensitivity of the SAF algorithm to these parameters is assessed in section III.

D. Working principle

The key principle of the SAF algorithm is its capability to adapt the image fusion process to both the image content and the amount of noise without changing its parameters S_{th} and α. For instance, Fig. 5(a) shows a frame from two synthetic sequences corresponding to the same scene but with different noise levels. The corresponding values of S and $\phi_{x,y}$ are shown in Fig. 5(b) and (c), respectively.

In Fig. 5, the first row corresponds to a low-noise focus sequence. In this case, the selectivity measure, S, is high all over the image, meaning that the pixel intensities and their corresponding focus measures are reliable. This leads to high values of $\phi_{x,y}$. Thus, the image fusion process is mostly performed through an energy maximization scheme. In contrast, the second row of Fig. 5 corresponds to a sequence with high levels of noise. The sharpening parameter $\phi_{x,y}$ has high values only for those pixels where image features are strong enough to compensate for the effects of noise. The low values of $\phi_{x,y}$ in areas where the image patterns are weak will lead to a stronger smoothing, hence suppressing noise.

The main conceptual difference between the proposed approach and previous works is that, instead of applying the same smoothing rule to the whole image (by means of low pass filters, Gaussian pyramids or by removing wavelet coefficients), the image fusion is performed adaptively by taking into account the local features of the scene and the response of the focus operators to those features. This leads to a reduction of both noise and artifacts while preserving image texture.

III. Experiments

Several experiments have been conducted in order to assess the performance of the proposed approach with either low noise or high noise for real and synthetic focus sequences. Details regarding the generation of the synthetic sequences are also provided, as well as on the acquisition of real focus sequences with varying levels of noise.

A. Defocus simulation

A defocused image is often considered in the literature as a filtered version of a focused one. Thus, a defocused image I_d can be described as the convolution of the focused image I with a blurring function h:

$$I_d = I * h. \tag{9}$$

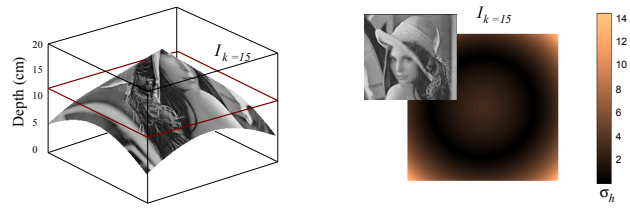
Function h is referred to as a *Point Spread Function* (PSF), since it is the response of the camera to a unit point source [21]. In diffraction limited optics with incoherent illumination, the PSF can be simplified as a Gaussian [21] with variance σ_h, which is assumed to be proportional to the degree of defocus of the image. Following [22], it is possible to derive an expression that interrelates the blur parameter σ_h, the distance between the imaged point and the in-focus position, δ, and fixed camera parameters such as the lens focal length f_L, the f-number N and the pixel size κ:

$$\sigma_h = \frac{\kappa f_L^2}{N} \frac{\delta}{u(u_f - f_L)}, \tag{10}$$

where u_f is the in-focus position (the distance at which the camera is focusing).

The convolution in (9) is only valid under the assumption of a spatially invariant blurring function within the evaluation window (isoplanatism). Therefore, in order to avoid the isoplanatism assumption for the synthetic data used in this work, a blurred image $B_{x,y}$ is obtained for every scene point at coordinates (x, y) by convolving it with its corresponding PSF: $B_{x,y} = I(x, y) * h_{x,y}$, where $h_{x,y}$ denotes the PSF corresponding to pixel $I(x, y)$ according to its depth. In turn, the defocused image for the pixel located at (x_0, y_0) given a $W \times H$ image is obtained by adding the contributions of every defocused point:

$$I_d(x_0, y_0) = \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} B_{i,j}(i - x_0, j - y_0). \tag{11}$$



(a) (b)

Fig. 6. Simulation of a focus sequence. (a) Synthetic textured surface. (b) Blur width of the 15-th frame.

Image noise must be taken into account for a more accurate simulation of the defocus process. A CCD camera has several primary noise sources, which can be grouped into irradiance-dependent and irradiance-independent sources. In that way, a noisy image I_n can be modeled as [23]: $I_n = f(I + n_s + n_c) + n_q$, where I is the original image, $f(\cdot)$ is the camera response function (CRF), n_s is the irradiance-dependent noise component, n_c is the independent noise, and n_q is the additional quantization and amplification noise. According to [23], n_q is neglected, n_s and n_c are assumed to have zero mean and variances Var(n_s) = $I \cdot \sigma_s^2$ and Var(n_c) = σ_c^2, respectively.

In this work, all simulated sequences consist of 30 images of different sizes. The textures mapped on the synthetic surfaces were selected in order to cope with a variety of features and intensities. For illustration purposes, Fig. 6 shows a surface with a texture mapped on it and the blur width of each pixel in one of the resulting simulated frames. The continuous shape of the conic surface in Fig. 6 yields frames with different degrees of blur ($0 < \sigma_h < 15$).

B. Tests on simulated data

The focus sequences synthetically generated as described in the previous section allow the availability of a ground truth for an objective estimation of the performance of the image fusion process. Based on a thorough review of the literature, five algorithms were selected for comparison:

1) *Helicon Focus*: an image fusion software produced by HeliconSoft [24].
2) *Zerene Stacker*: a fusion software produced by ZereneSystems [25].
3) Extended depth of field (EDF): a fusion algorithm based on wavelets presented in [7].
4) 3D extended depth of field (3D EDF): a fusion algorithm based on defocus modeling presented in [6].
5) The algorithm proposed by Tian *et al.* [26] based on the spatial frequency.

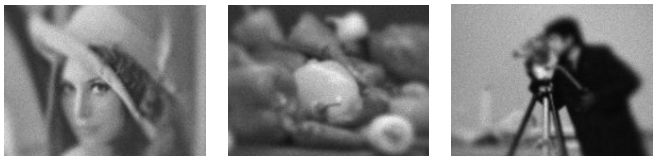


Fig. 7. Synthetic focus sequences used for comparison

Fig. 9. Image fusion for synthetic focus sequences obtained with different algorithms. From top to bottom: Zerene Stacker [25], Algorithm of Tian. *et al.* [26], EDF [7], Helicon Focus [24], 3D EDF [6] and proposed SAF algorithm.

Figure 8 shows the mean performance of the different algorithms in terms of signal-to-noise ratio: SNR = $20\log(\Sigma_{x,y}I(x,y)/\Sigma_{x,y}|I(x,y)-\phi(x,y)|)$, the peak-signal-to-noise ratio PSNR = $10\log(\Sigma_{x,y}255^2/\Sigma_{x,y}(I(x,y)-\phi(x,y))^2)$ and the *universal quality index* (UQI) originally proposed in [27]. The i-th noise level corresponds to noise variances $\sigma_c^2=\sigma_s^2=0.06i$. In this figure, *Ref* corresponds to the mean performance of the original focus sequence used to generate the AIF image with respect to the associated noiseless focus sequence. Fig. 7 shows a frame from three different synthetic focus sequences, whereas Fig. 9 shows details of the all-in-focus images obtained from the previous sequences using the evaluated algorithms. For high noise levels, the difference in quality of the AIF image obtained with the different methods tends to increase in favor of the proposed SAF algorithm. The images shown in Fig. 9 correspond to the first noise level ($i=1$)[1].

[1]A full resolution version of all the images shown in this work and the parameters of each stacking algorithm can be found online at http://www.sayonics.com/research/focus_fusion.html

TABLE I
NOISE LEVELS FOR REAL SEQUENCES

Noise level	Gain [dB]	Shutter speed [s]
0	0	1/12
1	+16	1/75
2	+20	1/120
3	+22	1/150
4	+26	1/215
5	+28	1/300

Fig. 10. Real scene at increasing noise levels. Left to right: 0th, 2nd and 5th noise level.

C. Tests on real data

Real focus sequences of 33 images of 640×480 pixels were acquired with a Sony SNC-RZ50P camera. In order to increase the noise level, the shutter speed was reduced. Since the intensity of an image pixel is proportional to the integration time (inversely related to the shutter speed), the loss of intensity is then compensated by the gain of the camera.[2] This leads to an increase of the *amplification noise*. The procedure for capturing real sequences with different noise levels can be summarized as follows:

1) The camera is adjusted to obtain the best-quality image of the scene and a focus sequence is captured. This sequence corresponds to noise level 0.
2) The gain of the camera is increased in order to raise the amplification noise. The shutter speed is increased in order to compensate for the illumination change. This sequence corresponds to noise level 1.
3) Since the camera gain and shutter speed can only be set to discrete predefined values, the histograms of the captured images must be equalized as necessary.
4) Steps 2 and 3 are repeated to compute the sequence corresponding to the i-th noise level.

Table I summarizes the camera configuration used for the acquisition of real sequences at different noise levels. Fig. 10 shows a frame of a particular scene with different noise levels.

Tests have been conducted on both color and gray scale images. For the color images, each frame was converted to gray scale in order to compute both the focus measure (1) and the selectivity measure (4). Then, the fusion rule in (5) was independently applied to each color channel.

For the real sequences, it is not possible to compute an objective quantitative performance measure since the ground-truth is not available and the quality of the results must be subjectively determined by simple observation. However, it is possible to assess the impact of noise over the fusion process by comparing the AIF image obtained from a noisy sequence against the AIF image obtained from the sequence with the

[2]The lens aperture must remain unchanged in order to keep the same depth of field.

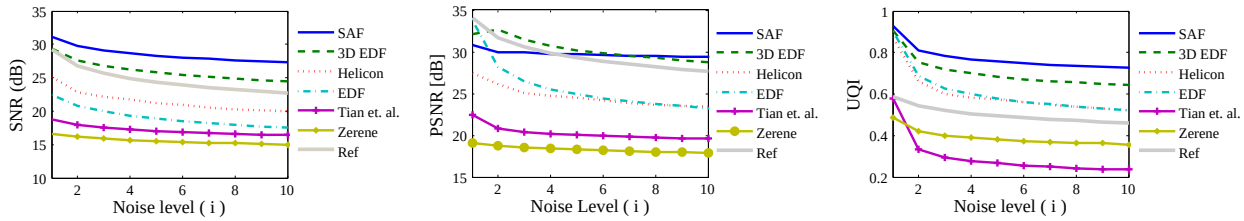


Fig. 8. Performance comparison for synthetic focus sequences.

Fig. 12. Image fusion for a real focus sequence obtained with different algorithms. From top to down: first row, Helicon Focus software [24]; second row, 3D EDF algorithm [6]; third row, proposed SAF algorithm; fourth row, $\phi_{x,y}$.

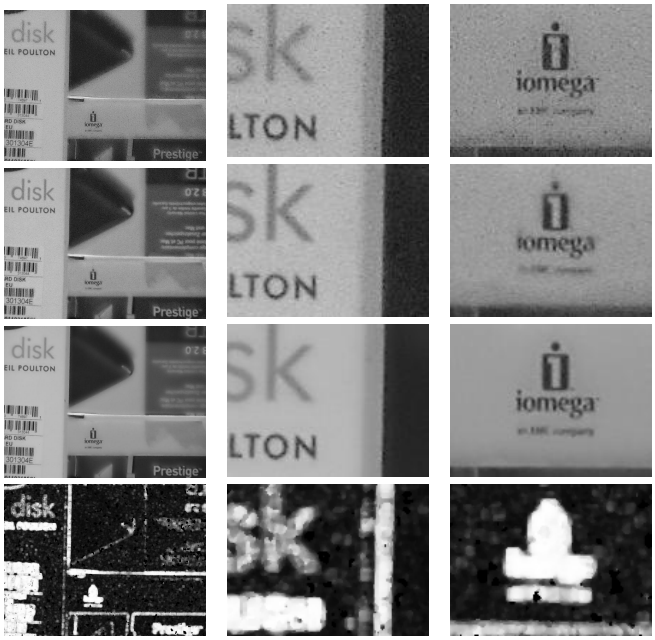


Fig. 13. Image fusion for a gray-scale real focus sequence with different algorithms. From top to down: first row, Helicon Focus software [24]; second row, 3D EDF algorithm [6]; third row, Proposed SAF algorithm; fourth row, $\phi_{x,y}$.

lowest noise level. Thus, in Fig. 11, the SNR is computed using the all-in-focus image obtained from the sequence with the lowest noise level as a reference. For an algorithm to be robust to noise, the AIF images of sequences with higher noise will be less corrupted and will, therefore, have a higher SNR. Fig. 12 shows details of the all-in-focus images obtained from a color focus sequence using Helicon Focus, 3D EDF and SAF. In addition, this figure shows the sharpening parameter, $\phi_{x,y}$, in order to illustrate how the fusion process is performed. Fig. 13 is an example for a gray-scale sequence.

D. Algorithm's performance

As shown in section II, the proposed algorithm depends on two parameters. The values of the parameters used in the results shown in this work for both synthetic and real sequences correspond to $S_{th} = 11$ dB and $\alpha = 0.2$. These parameters have been selected experimentally using the synthetic sequences and recording the mean SNR for each parameter pair.

As shown in Fig. 14, the parameters were selected in order to maximize the performance in the synthetic sequences

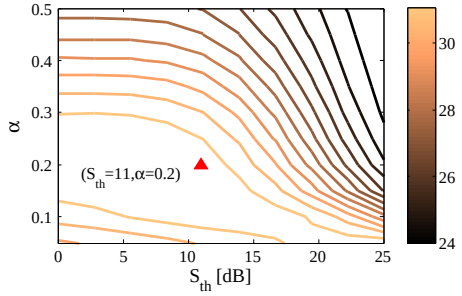


Fig. 14. Level plot of the mean SNR of synthetic sequences as a function of the parameters α and S_{th}. The best performance is obtained for $S_{th} = 11$ and $\alpha = 0.2$.

(in terms of SNR). In order to assess the sensitivity of the proposed approach to these parameters, the synthetic focus sequences corresponding to noise level 1 were processed with variations of those parameters of $\pm 15\%$. The maximum variation observed in the SNR was -1 dB (3.1%). With this variation, the SAF algorithm still outperforms the closest competing algorithm (3D EDF). The variation of SNR for real

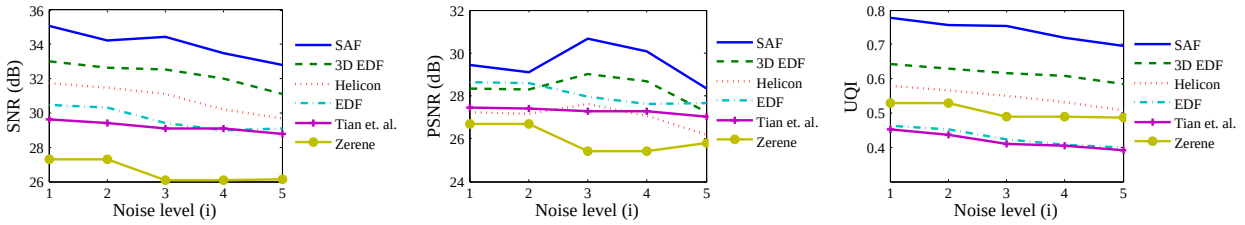


Fig. 11. Performance comparison for real focus sequences.

Fig. 15. Three frames from real sequence ($f_L = 42$ mm, $N = 1.6$). Left to right: 1st, 10th and 30th frame. The color rectangles in the first image highlight the first (red), second (green) and third (blue) region of interest.

sequences was 1.5%.

According to (10), the factors that most influence the amount of blur (and hence the performance of any focus measure operator) are the parameters of the acquisition process.

Although it is difficult to evaluate the performance of image fusion algorithms under all possible imaging conditions, the following experiment was conducted in order to identify the limitations of the proposed approach: Fig. 15 shows three frames of a real focus sequence. For this particular scene, a focal length of $f_L = 42$ mm was used and the different objects placed in order to cover a long distance range (between 0.5 m and 3.5 m from the camera). According to (10), the object position u influences the shape of the focus function. Fig. 15(a) highlights three different regions of interest corresponding to objects at different depths. In addition, the same scene was captured using three different f-numbers.

The response of the selective all-in-focus to different object distances and lens f-number is illustrated in Fig. 16. In Fig. 16(a), the weighting ω (low row) has a high slope and is the least sensitive to the change in f-number. This behavior is sound since the focus function (top row) has a sharp peak around the in-focus position. Therefore, the all-in-focus pixels should be constructed by giving a high weight to the pixels with high focus value in the original sequence. The focus functions in Fig. 16(b) (top row) show that the corresponding object is rapidly focused (frames 1 to 5) but it defocuses very slowly, leading to a flat focus profile. This yields a weighting function with a low slope that composes the all-in-focus image using more images from the original sequence. Notwithstanding in this particular case, the contrast is not affected since many images of the original sequence have a high focus value (from the fifth frame to the last). Finally in Fig. 16(c), the focus function increases slowly and the weighting ω is the most sensitive to changes in the f-number.

Fig. 17 shows the AIF images and some details for sequences of the same scene shown in Fig. 15 captured with different f-numbers. Since the f-number changes the amount

Fig. 17. Details of image fusion for objects with different focus functions and different f-number. From left to right: f-number = 1.6, 4.0 and 8.0, respectively.

of light that reaches the camera's sensor, the change in illumination has been compensated by changing either the shutter speed or the gain. Therefore, it is possible to observe differences in the image balance.

In the second and last rows of Fig. 17, it is possible to observe some areas with low contrast mainly due to two reasons: first, the composition of images with high variations of defocus may be deficient if the weighting function ω is not sharp enough and fails to select only the pixels with highest focus value in regions with rich texture. Secondly, the Gaussian fit used to model the peak of the focus function adapts to the image content poorly when the depth-of-field is large and the focus variation slow.

IV. Discussion

The results presented in section III show that the proposed method outperforms the other tested algorithms for synthetic sequences even at the lowest noise levels ($i = 1$). For instance, in the first column of Fig. 9, the face has a smoother appearance, whereas highly-textured areas, such as the hat's feathers, are sharply recovered. In the image of the cameraman, the sky shows a cleaner appearance, whereas the contours of the man are well defined.

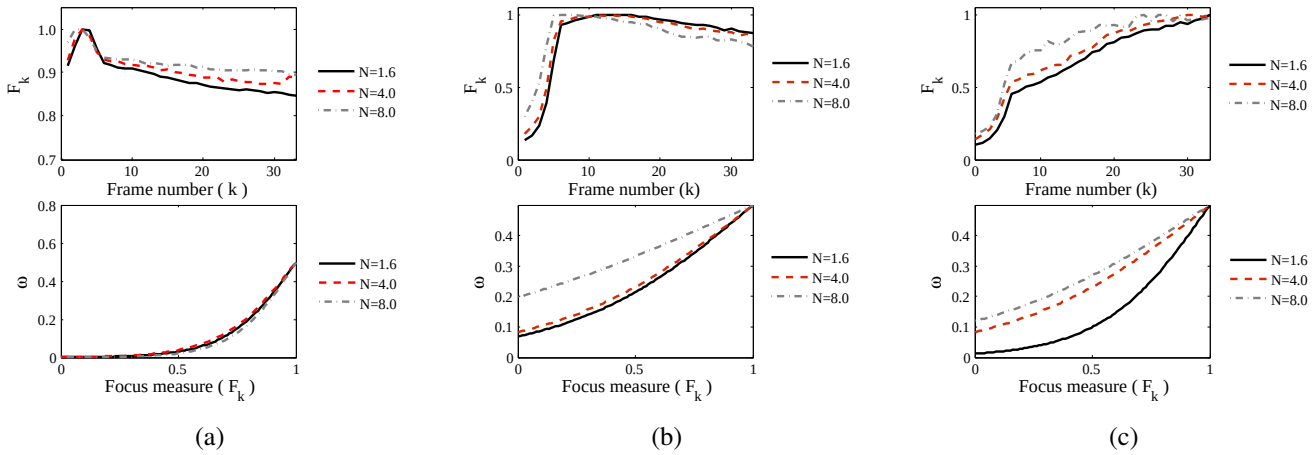


Fig. 16. Selective weighting for different focus profiles. First row: focus functions. Second row: weighting function. First region of interest (a), Second region of interest (b) and third region of interest (c) of Fig. 15.

In real sequences, the results show that the SAF algorithm is the most insensitive to noise, confirming the results obtained with synthetic sequences. For example, in the detail images of Fig. 12 and 13, the background always presents a cleaner appearance, whereas edges, letters and contours are sharply defined even for thin and small characters. However, a white halo can be observed near the edges of the white letters within the detail image in Fig. 12. This effect is barely noticeable and is observed at higher noise levels around bright white areas surrounded by a dark background. This halo can be due to the fact that bright spots have a larger spread than dark ones when defocusing, and the radiance of these spots may "leak" into darker areas during the image fusion process.

From the results obtained for a same scene with different noise levels, it is evident that the proposed approach responds to the image content selectively. Therefore, as the noise level increases, the fusion process provides a smooth low-noise response in areas of low PSNR, while reducing the negative impact on image features. From the results obtained using non-Gaussian focus profiles, the proposed approach yielded acceptable results. Notwithstanding, the results may be improved with a more accurate model for the focus profile.

The need for parameters is common in the evaluated AIF algorithms. Notwithstanding, the results in the previous section show that the proposed method is reasonably insensitive to its parameterization. The fact that the transfer function of (7) and (8) is defined in terms of sigmoids guarantees that the intensity of a given pixel of the AIF image will always be a combination of the pixel intensities of the focus sequence. This has a positive impact on the stability of the algorithm.

In terms of computational cost, the least complex all-in-focus methods are those based on the spatial frequency. These methods usually imply the application of a focus measure to each image of the focus sequence, followed by a fusion rule. In the second place, the pyramid-based and wavelet-based approaches usually require a forward transform, a combination step applied to the obtained sub-images or sub-bands, and an inverse transform. The cost of the forward and inverse transforms increases with the number of levels of the pyramid. The methods with the highest computational costs are those based on defocus modeling (e.g., 3D EDF).

The different methods compared in this work were obtained from different sources and platforms (e.g., Java, Matlab, C). Therefore, an objective quantitative comparison in terms of computation time is not provided. Notwithstanding, the efficiency of the proposed approach is between that of spatial-based methods and wavelet-based methods. Similarly to spatial-based methods, the SAF algorithm requires the application of a focus measure followed by a fusion rule. However, the computation of the selectivity measure represents an additional cost. In spite of that, the computation of both the focus measure (1) and the selectivity measure (4) is simple and fast. In particular, the computational complexity of the proposed algorithm is O(NKh^2), where K is the number of images, N the number of pixels in each image and h the radius in (1). The Matlab implementation of the proposed algorithm fuses a sequence of 50 gray-scale images of 640×480 pixels in approximately 7.0 s running on an Intel 2 Quad processor at 2.5 GHz and 4GB of RAM.

V. Conclusions

A new method to generate the all-in-focus image from a focus sequence has been presented. The proposed approach selectively fuses the different frames of the focus sequence in order to reduce noise while preserving image features. The selective fusion is performed in three steps: focus measure, selectivity measure and image fusion. The proposed algorithm has been compared to state-of-the-art methods upon both synthetic and real focus sequences. A Matlab toolbox for the computation of the all-in-focus image using the SAF algorithm has been implemented and will be available online. The proposed method has been applied to gray-scale images and extended to color images by applying the fusion rule independently to each color plane.

The performance of the algorithm depends on the model selected for the focus profile. Future work will compare different models, such as those proposed in [17] and [18]. In addition, the final results are expected to improve if the

algorithm is applied with a pre-processing that compensates for the image shift observed in focus sequences.

References

[1] J. Tian and L. Chen, "Multi-focus image fusion using wavelet-domain statistics," in *proc. IEEE Int. Conf. Image Process.*, 2010, pp. 1205 – 1208.

[2] M. Antunes, M. Trachtenberg, G. Thomas, and T. Shoa, "All-in-focus imaging using a series of images on different focal planes," in *Image Analysis and Recognition*, 2005, vol. 3656, pp. 174–181.

[3] Z. Zhang and R. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proc. IEEE*, vol. 87, no. 8, pp. 1315 –1326, 1999.

[4] M. Subbarao and T. Choi, "Accurate recovery of three-dimensional shape from image focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 3, pp. 266–274, Mar 1995.

[5] K. Kodama, H. Mo, and A. Kubota, "Simple and fast all-in-focus image reconstruction based on three-dimensional/two-dimensional transform and filtering," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1, 2007, pp. 769 –772.

[6] F. Aguet, D. Van De Ville, and M. Unser, "Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy," *IEEE Trans. Image Process.*, vol. 17, no. 7, p. 1144–1153, July 2008.

[7] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser, "Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images," *Microsc. Res. Tech.*, vol. 65, no. 1-2, pp. 33–42, September 2004.

[8] M. Born and E. Wolf, *Principles of Optics*, 7th ed. Cambridge University Press, 1999.

[9] M. Subbarao and J.-K. Tian, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 864–870, August 1998.

[10] S. Li, J. T. Kwok, and Y. Wang, "Combination of images with diverse focuses using the spatial frequency," *Information Fusion*, vol. 2, no. 3, pp. 169 – 176, 2001.

[11] V. Aslantas and R. Kurban, "A comparison of criterion functions for fusion of multi-focus noisy images," *Opt. Commun.*, vol. 282, no. 16, p. 3231–3242, 2009.

[12] A. G. Valdecasas, D. Marshall, J. M. Becerra, and J. J. Terrero, "On the extended depth of focus algorithms for bright field microscopy," *Micron*, vol. 32, no. 6, pp. 559 – 569, 2001.

[13] Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microsc. Res. Tech.*, vol. 65, no. 3, pp. 139–149, 2004.

[14] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug 1994.

[15] J. Lee, K. Kim, B. Nam, J. Lee, Y. Kwon, and H. Kim, "Implementation of a passive automatic focusing algorithm for digital still camera," *IEEE Trans. Consumer Electronics*, vol. 41, no. 3, pp. 449–454, Aug 1995.

[16] A. S. Malik and T.-S. Choi, "Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3d shape recovery," *Pattern Recognit.*, vol. 40, no. 1, pp. 154–170, 2007.

[17] M. Muhammad and T.-S. Choi, "Sampling for shape from focus in optical microscopy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 564 –573, mar 2012.

[18] D. Tsai and H. Chen, "Reciprocal focus profile," *IEEE. Trans. Image Process.*, vol. 21, no. 2, pp. 459 –468, feb. 2012.

[19] G. Wolberg, *Digital Image Warping*. IEEE Comput. Soc. Press, 1990.

[20] A. Basokur, "Digital filter design using hyperbolic tangent functions," *J. Balkan geophys. soc.*, vol. 1, pp. 14–18, 1998.

[21] M. Subbarao, T. Choi, and A. Nikzad, "Focusing techniques," *Opt. Eng.*, vol. 32, pp. 2824–2836, 1993.

[22] A. P. Pentland, "A new sense for depth of field," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 4, pp. 523–531, July 1987.

[23] C. Liu, R. Szeliski, S. B. Kang, C. Zitnick, and W. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 299 –314, feb. 2008.

[24] Helicon Soft, "Helicon focus," retrieved on 01/04/2011. [Online]. Available: http://www.heliconsoft.com/heliconfocus.html

[25] Zerene Systems, "Zerene stacker," retrieved on 01/04/2011. [Online]. Available: http://www.zerenesystems.com/cms/stacker

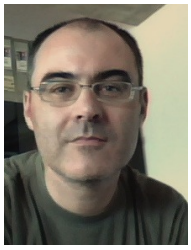
[26] J. Tian, L. Chen, L. Ma, and W. Yu, "Multi-focus image fusion using a bilateral gradient-based sharpness criterion," *Opt. Commun.*, vol. 284, no. 1, pp. 80 – 87, 2011.

[27] Z. Wang and A. Bovik, "A universal image quality index," *Signal Processing Letters, IEEE*, vol. 9, no. 3, pp. 81 –84, mar 2002.

Said Pertuz received the B.S. degree in Electronics Engineering from Industrial University of Santander, Colombia (2007) and the M.S. degree in Computer Science from Rovira i Virgily University, Spain (2009). In 2009, he joined the Intelligent Robotics and Computer Vision group at Rovira Virgili University as Ph.D student. His current research interests include focus-related applications in conventional cameras, such as autofocus, focus stacking, focus measure, depth-of-field modeling and shape from focus.

Domenec Puig received the M.S. and Ph.D. degrees in computer science from Polytechnic University of Catalonia, Barcelona, Spain, in 1992 and 2004 respectively. In 1992, he joined the Department of Computer Science and Mathematics at Rovira i Virgili University, Tarragona, Spain, where he is currently Associate Professor. Since July 2006, he is the Head of the Intelligent Robotics and Computer Vision group at the same university. His research interests include image processing, texture analysis, perceptual models for image analysis, scene analysis, and mobile robotics.

Miguel Angel Garcia received the B.S., M.S., and Ph.D. degrees in computer science from Polytechnic University of Catalonia, Barcelona, Spain, in 1989, 1991, and 1996 respectively. He joined the Department of Software at Polytechnic University of Catalonia in 1996 as an Assistant Professor. From 1997 to 2006, he was with the Department of Computer Science and Mathematics at Rovira i Virgili University, Tarragona, Spain, where he was the Head of Intelligent Robotics and Computer Vision group. In 2006, he joined the Autonomous University of Madrid, Spain, where he is currently Associate Professor. His research interests include mobile robotics, image processing, and 3D modeling.



Andrea Fusiello received his Laurea (M.S.) degree in Computer Science from the University of Udine in 1994. He received the Dottorato di Ricerca (PhD) in Computer Engineering from the University of Trieste in 1999. He had been Visiting Research Fellow at Heriot-Watt University, Edinburgh in 1999 supported by an EPSRC grant. From 2001 to 2004 he served as a Ricercatore (Assistant Professor) at the Department of Computer Science, University of Verona, and from 2005 to 2011 as Professore Associato (Associate Professor) at the same department. In 2012 he joined the Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica, University of Udine, where he teaches Computer Vision and Computer Science basics. His research is mainly focused on Computer Vision: image analysis, 3D model acquisition, image based rendering.