Repositorio Institucional de la Universidad Autónoma de Madrid

https://repositorio.uam.es

Esta es la **versión de autor** de la comunicación de congreso publicada en:
This is an **author produced version** of a paper published in:

Journal Image and Vision Computing 31.10 (2013): 725-734

DOI: http://dx.doi.org/10.1016/j.imavis.2013.07.005

Copyright: © 2013 Elsevier B.V. All rights reserved

Reliability measure for shape-from-focus

Said Pertuz[a], Domenec Puig[a], Miguel Angel Garcia[b]

[a]*Intelligent Robotics and Computer Vision Group*
Department of Computer Science and Mathematics.
Universitat Rovira i Virgili.
[b]*Department of Electronic and Communications Technology*
Autonomous University of Madrid.

Abstract

Shape-from-focus (SFF) is a passive technique widely used in image processing for obtaining depth-maps. This technique is attractive since it only requires a single monocular camera with focus control, thus avoiding correspondence problems typically found in stereo, as well as more expensive capturing devices. However, one of its main drawbacks is its poor performance when the change in the focus level is difficult to detect. Most research in SFF has focused on improving the accuracy of the depth estimation. Less attention has been paid to the problem of providing quality measures in order to predict the performance of SFF without prior knowledge of the recovered scene. This paper proposes a reliability measure aimed at assessing the quality of the depth-map obtained using SFF. The proposed reliability measure (the *R-measure*) analyses the shape of the focus measure function and estimates the likelihood of obtaining an accurate depth estimation without any previous knowledge of the recovered scene. The proposed *R*-measure is then applied for determining the image regions where SFF will not perform correctly in order to discard them. Experiments with both synthetic and real scenes are presented.

Keywords: Image sequences, focus measure, shape from focus, reliability, detph-map carving

1. Introduction

A great effort has been devoted in the field of image processing in order to develop and improve both active and passive techniques for depth recovery and 3D shape reconstruction. In particular, *shape-from-focus* (SFF) is a passive technique for depth recovery that requires a sequence of images from a scene captured by changing the focus configuration of the imaging device. This image sequence is usually referred to as *focus sequence* or *focus sweep*. The main interest of SFF is that, as a passive method, there is no physical interaction between the capturing device and the scene. In addition, since a single camera is used, correspondence problems typical of stereo or vergence are avoided. Moreover, the hardware complexity is low, with no special equipment being required, as in the case of active techniques such as optical triangulation, LIDAR imaging and laser scanning.

An important step in the application of SFF is the computation of the focus level for every scene point in every image frame of the focusing sequence. Many *focus measure* (FM) operators have been proposed and tested in the literature. All of them require that the change in the focus level of the imaged points be detectable. This is an important limitation if SFF is to be applied to complex scenes captured with large depth-of-field (DOF) systems where the optical configuration of the capturing

device and the image content may lead to erroneous detections of the change in focus. In this work, the term *"large DOF systems"* refers to conventional macroscopic cameras in which the DOF is large with respect to the working distance. In contrast, in microscopy imaging, the large focal lengths lead to very short DOFs; and the captured images are usually rich in texture due to the high magnification. In addition, the focus sweep is obtained by moving the captured object while the camera's optics remains fixed, yielding an acquisition with constant DOF. For this reason, SFF has typically been used in microscopy and in well-controlled scenarios, where favorable imaging conditions are guaranteed prior to the application of that technique. For instance, SFF has been applied in microscopy for PCB inspection and manufacturing [1].

Due to the aforementioned reasons, the application of pure SFF in conventional cameras has been limited. For instance, in [2] and [3], a depth-from-defocus prototype that projects a light pattern in order to compensate for the lack of texture was developed. In [4], a hybrid system that combines SFF and stereo has been proposed. In [5, 6], the result of SFF is improved using the information of relative defocus.

The problem of applying SFF to complex scenes using conventional cameras is tackled in this work. With the proposed methodology, we show that even in complex scenarios, useful information about the scene depth can be obtained by means of SFF as long as a way for measuring the reliability of the depth estimation is available. For instance, Fig. 1(a) shows a synthetic scene that consists of a conical surface with a texture mapped on it. Fig. 1(b) shows the corresponding depth-map obtained
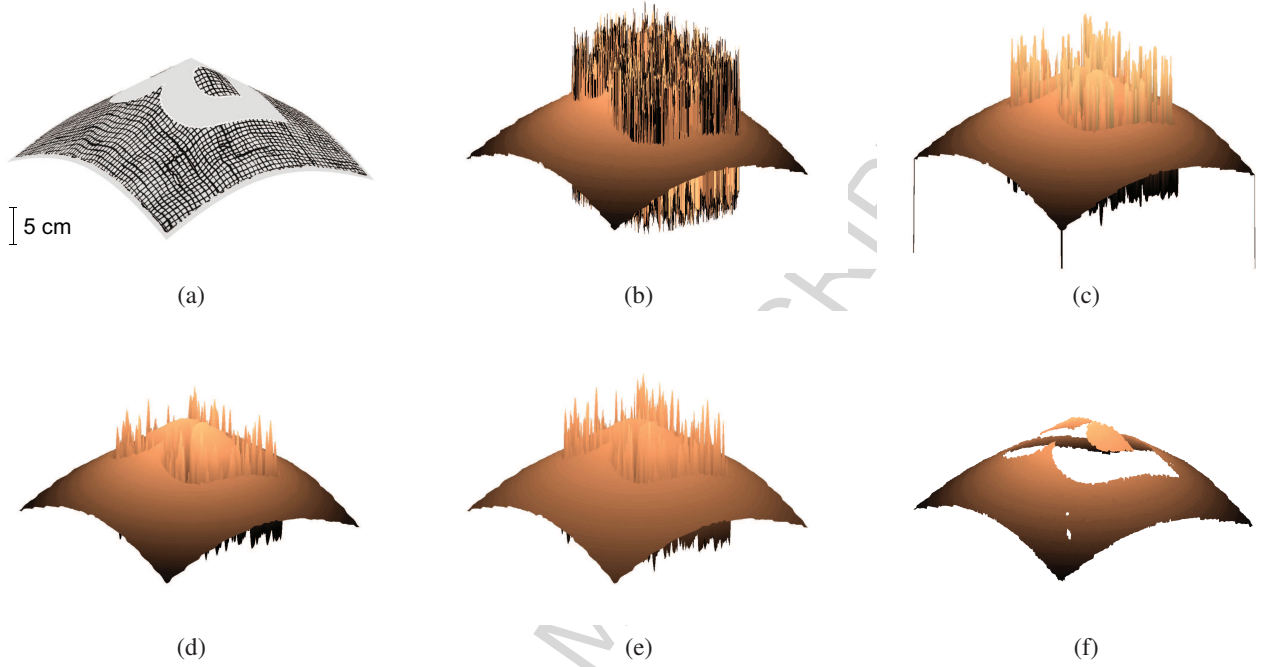
Fig. 1: Shape-from-focus in complex scenarios. (a) Synthetic scene. (b) Depth-map obtained through SFF. (c) Smoothing with median filter [7, 8]. (d) Smoothing with bilateral filter [9]. (e) Smoothing with non-local means [10, 11]. (f) Carved depth-map with low-reliability pixels having been removed.

through SFF. It is clear that some regions of the obtained reconstruction are inaccurate and highly corrupted by noise. In this case, traditional smoothing techniques such as median filtering [7, 8], bilateral filtering [9] or non-local means [10, 11] are of limited application since large areas of the recovered scene are unreliable. In contrast, Fig. 1(f) shows the depth-map of Fig. 1(b) after having been filtered by carving those pixels whose reliability is below a given threshold. The reliability measure allows identifying and removing less accurate pixels while preserving the useful information of the depth-map. In contrast to previous approaches, the reliability of each pixel is estimated before the computation of the depth-map based on the behavior of the focus measure over each pixel of the imaged scene. Experimental results on real and synthetic data are provided.

The problem of estimating the reliability of the focus measure is analogous to confidence estimation in stereo and optical flow [12, 13]. In that scope, the aim is to rank depth estimates in stereo vision or flow fields in optical flow according to the likelihood for being correct. To the best of our knowledge, the problem of determining the confidence of the focus measure estimation is tackled in this paper for the first time.

This paper is organized as follows: section 2 reviews the main concepts behind SFF and the relevant related previous work. Section 3 describes the proposed approach. Section 4 presents the experiments carried out on real and synthetic data, and discusses the obtained results. Finally, conclusions and future work are given in section 6.

2. Background

In SFF, depth information is retrieved from image sequences of the same scene captured with different degrees of focus. The local focus variation is used as a depth cue [7]. According to the *thin lens* camera model, a camera focuses by changing the distance, v, between the lens and the sensing device (e.g., a CCD sensor). Depending on the lens focal length, f, only the points at a certain distance u will be in focus for a given focus setting. The relationship between these three variables is given by the well known thin lens equation: $1/f = 1/u + 1/v$. Since there is a one-to-one correspondence between the object distance u and the focal plane location v, the maximum focus will only be achieved at a specific object distance.

The SFF problem can be divided into two main subproblems, namely focus measure and scene reconstruction. Once a focus sequence has been captured by changing focus and an image stack $I_k(x, y)$ is generated, a focus measure is computed for each pixel at every image frame. The values of the focus measure for a pixel at coordinates (i, j) over all the image frames are referred to as *focus function* (or focus measure vector): $f_{i,j} = (F_1(i, j), ...F_k(i, j), ...F_K(i, j))$, where $F_k(i, j)$ is the focus measure of that pixel at the k-th frame and K is the total number of frames.

The next step for shape reconstruction is to apply a scene reconstruction scheme that exploits the focus information of images in order to estimate the distance of every point of the scene, namely the *depth-map*, $Z(x, y)$. The reconstruction stage usually requires the interpolation of the focus function by means of a particular model [8, 14, 15]. The final reconstruction can

2

be further improved by applying optimization techniques that post-process the focus measure and the reconstructed surface [14, 16, 17] or exploit the defocus information [5].

The influence of the image content on the performance of SFF has been analyzed in the literature. In [18], the presence of texture content in the focus sequence was found to be critical, since the focus measure operators fail to detect changes in focus for low-textured scenes, thus making the estimation of depth unreliable. To our knowledge, few attempts have been made in order to address the problem of detecting regions with low accuracy in SFF. In [19], Shoji *et. al.* used color segmentation and bilateral filtering to improve both the accuracy of the focus measurement and the final estimation of depth. The main drawback of this approach is that the region merging stage is performed by taking into account color and not texture, which is the key factor for the focus measure. This may lead to erroneous results for different objects and surfaces with the same color. Both [18] and [19] do not provide information about the distribution of reliable points in the depth-map.

In [20], a depth-map is initially obtained using traditional SFF. Then, parts of the scene with high depth variations are discarded by assuming that they are due to an inaccurate computation of the focus measure. The discarded regions are then recovered by interpolation. The disadvantage of the latter approach is that it is only applicable to scenes where non-reliable regions are small and can be interpolated from highly-textured ones. In [21], Gaganov and Ignatenko apply Markov random fields in order to smooth the obtained depth-map in low-reliability areas. This approach also assumes that the depth information of highly-textured areas can be used to infer and constrain the depth-map where SFF fails, but it does not provide information about the location of low- and high-reliability areas. More recently, Muhammad and Choi [22] have proposed to carve the depth-map by applying a Canny edge detector to the all-in-focus image of the scene.

The main contribution of the present work is a technique that aims at measuring the reliability of SFF according to the behavior of the focus measure. In contrast to previous approaches, the proposed methodology computes the reliability measure without the need for either computing the all-in-focus image or post-processing the generated depth-map. In this sense, the reliability measure aims at assessing the confidence on the performance of SFF. The proposed approach efficiently integrates with the SFF framework. The experimental tests presented in this paper show that this method is able to accurately detect non-reliable regions in terms of SFF.

3. Methodology

This section describes an algorithm to determine the reliability of SFF. In order to illustrate its applicability, the proposed reliability measure is utilized for determining the image regions where the depth estimation using SFF is unreliable in order to remove them from the depth-map. This task is performed in two main stages that take place after the application of a focus measure to the image stack: in the first stage, the reliability measure (*R measure*) is assigned to the depth estimation of every pixel. This reliability is based on the behavior of the focus function. In the second stage, the *R* measure is used to carve the obtained depth-map by applying a threshold obtained as a result of a training process. The *R* measure and the carving stages are described in further detail below.

3.1. Focus measure

In order to obtain a depth-map with SFF, it is first necessary to compute the focus measure (FM) for every pixel of the image stack. Several algorithms and operators have been proposed for this purpose. The most popular operators are the Modified Laplacian [8], the Tenengrad Algorithm [23] and the Gray-Level Variance [20], among others [24]. In the present work, the Modified Laplacian is used as the focus measure:

$$F_k = \Delta_m I * h_r, \tag{1}$$

where h_r is a mean filter mask with radius r and $*$ denotes convolution. The size of r, also referred to as the evaluation window in the SFF literature, is application dependent and usually implies a tradeoff between robustness to noise and accuracy. A radius of $r = 9$ has experimentally been set in this work. $\Delta_m I$ is the modified Laplacian of I, computed as:

$$\Delta_m I = |I * \mathcal{L}_x| + |I * \mathcal{L}_y| \tag{2}$$

The convolution masks used to compute the Laplacian are $\mathcal{L}_x = [-1, 2, -1]$ and $\mathcal{L}_y = \mathcal{L}_x^T$.

The methodology presented in the next section can be readily adapted to deal with any of the focus measures operators found in the literature. A detailed study of focus measure operators in shape-from-focus can be found in [25].

3.2. Reliability measure

The shape of the focus function and, hence, the quality of the depth-maps obtained from SFF, depends on two main factors: the optical configuration of the capturing device and the image content. Due to the dependence on both factors and their interaction, the focus measure operator will respond differently when applied to the focus stack. This will determine whether the reconstruction technique will yield a successful depth-map estimation or not.

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

$$I_d = I * h. \tag{3}$$

Function h is referred to as a *Point Spread Function* (PSF), since it is the response of the camera to a unit point source [26]. In diffraction limited optics with incoherent illumination, the PSF can be simplified as a Gaussian [26, 27, 28] with variance σ_h, which is assumed to be proportional to the degree of defocus of the image. In turn, σ_h depends on physical variables of the acquisition device, such as the lens focal length, aperture and pixel size.
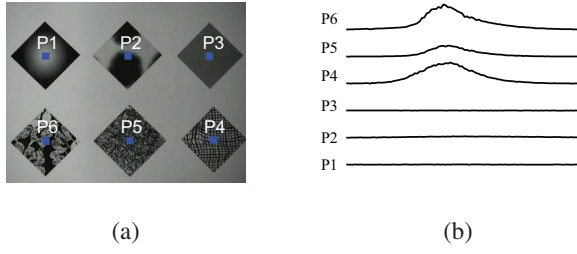
Fig. 2: Behavior of focus measures for different textures. (a) Selected points. (b) Corresponding focus functions

In addition to the optical configuration of the acquisition device, the image content directly affects the response of the FM operators. Most FM operators rely on their ability to detect high frequencies or spatial variations of the image in order to compute the degree of focus. This makes FM operators sensitive to image characteristics such as gray-level intensities, contrast and noise level. For illustration, Fig. 2 shows the focus measure functions corresponding to different textured patterns.

As can be appreciated in Fig. 2, the image content mostly determines the *strength* of the response of the focus measure operator. A reliability measure must be able to account for the variations of the focus function due to the configuration of the capturing device and the imaging conditions, as well as to detect when the focus operators fail to detect the change in focus. Bearing this in mind, the following methodology for computing the reliability measure is proposed.

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 an additive error signal:

$$f_{i,j} = G_{i,j} + E_{i,j}, \qquad (4)$$

where $f_{i,j}$ is the computed focus function for that pixel, $G_{i,j}$ an associated ideal focus function and $E_{i,j}$ an error signal that represents the departure of the focus function from the ideal behavior. $E_{i,j}$ accounts for the image noise, lack of texture, limitations of the focus measure operator or departure of the focus function from its ideal behavior.

Based on the empirical behavior of the focus function, some researchers have proposed to model $G_{i,j}$ as a Gaussian function. In this way, the idealized focus function corresponds to a smooth bell-shaped peak whose maximum corresponds to the position of the best focus. The real shape of the focus function in conventional cameras is still an open problem for both SFF and autofocus applications. In addition to the Gaussian model, quadratic and polynomial fits have also been used in order to approximate the focus function [14]. More recently, Tsai and Chen [29] and Mannan and Choi [22, 15] proposed a Lorentzian-Cauchy fit for the focus function in microscopy.

A discussion about the most accurate model for the focus measure function or a comparison of different models is beyond the scope of this work. However, despite the model chosen for the focus function, the performance of SFF depends on how accurately the focus function model in (4), $G_{i,j}$, approximates the measured focus function $f_{i,j}$. In the sequel, a reliability measure, $R1$, is first derived for the Gaussian model and, then, a general reliability measure, $R2$, compatible with any model for the focus function is presented.

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

$$G_{i,j}(z) = A \exp\left(\frac{-(z - \mu)^2}{2\sigma^2}\right), \qquad (5)$$

where A is the maximum value of the Gaussian function, μ its mean value and σ its standard deviation. Following [7], A, μ and σ are found by interpolation as a function of the depth value z.

In SFF, the estimated depth using the Gaussian model corresponds to the location of the maximum of $G_{i,j}$ (i. e., $z = \mu$). As claimed above, the estimated depth is likely to be reliable as long as there is a good fit between the measured focus function and the model. Inspired by the two-sample Kolmogorov-Smirnov test, the reliability measure for a pixel at coordinates (i, j) using the Gaussian model is obtained as:

$$R1(i, j) = 1 - \max\{|F(x|G) - F(x|f)|\}, \qquad (6)$$

where $\max\{\cdot\}$ denotes the *supremum* operator and $F(x|G)$ and $F(x|f)$ are the cumulative density functions (CDF) corresponding to $G_{i,j}$ and $f_{i,j}$, respectively. In (6), $G_{i,j}$ and $f_{i,j}$ are normalized so that $\sum_k G(k)_{i,j} = 1$ and $\sum_k f(k)_{i,j} = 1$. In this way, the measured focus function and the fitted Gaussian model are interpreted as probability density functions. In (6), the term $\max\{|F(x|G) - F(x|f)|\}$ is referred to as the Kolmogorov-Smirnov statistic. It measures the goodness of fit between two normal distributions. The CDFs in (6) are defined as:

$$F(x|f) = \sum_{k=1}^{x} f_{i,j}(k) \qquad (7)$$

$$F(x|G) = \frac{1}{2} + \frac{1}{2}\text{erf}\left(\frac{x - \mu}{\sqrt{2}\sigma}\right), \qquad (8)$$

where $\text{erf}(\cdot)$ is the so-called error function defined as:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \qquad (9)$$

In (7), $F(x|f)$ is referred to as the empirical cumulative density function. Equation (6) determines how well the measured focus function conforms to the desired behavior (a noiseless Gaussian-like peak) in the range [0,1]. The highest value 1 is achieved when the real data and the model match perfectly. Whenever the focus function departs from the idealized model due to the presence of noise, the image conditions or limitations in the focus measure operator, this reliability measure responds accordingly.

The advantages of R1 are the following: on the one hand, it exploits the information of the focus measure function. On the other hand, the Kolmogorov-Smirnov statistic is a well known and widely used tool for comparing probability distributions.
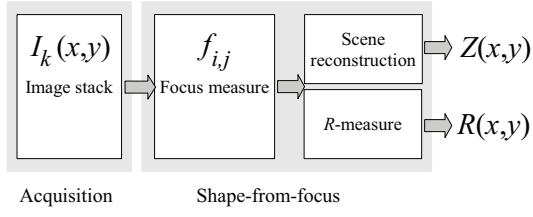
Fig. 3: R-measure and shape-from-focus pipeline. The R-measure efficiently integrates into SFF by exploiting the focus measure information.

Notwithstanding, the R-measure of (6) is limited to the Gaussian model. Alternatively, a more general reliability measure is defined as:

$$R2(i, j)^{-1} = \frac{1}{f_{max}K} \sum_{k=1}^{K} |f_{i,j}(k) - G_{i,j}(k)| \qquad (10)$$

$$R2(i, j)^{-1} = \frac{1}{f_{max}K} \sum_{k=1}^{K} |E_{i,j}(k)|, \qquad (11)$$

where $f_{max} = \max\{f_{i,j}\}$ is a normalization factor.

For convenience, let:

$$e_{i,j} = \frac{1}{K} \sum_{k=1}^{K} |E_{i,j}(k)|, \qquad (12)$$

denote the absolute average of the error signal. In this way, (11) is equivalent to (10), replacing the Kolmogorov-Smirnov statistic by $e_{i,j}$, which also accounts for the goodness of fit between the measured and the modeled focus function. The normalization factor f_{max} is necessary in order to guarantee that $R2(i, j) \geq 0$. The reliability measure in decibels is expressed as:

$$R2(i, j) = 20 \log \left(\frac{f_{max}}{e_{i,j}} \right), \qquad (13)$$

The aim of the logarithm in (13) is expressing R2 analogously to the peak signal-to-noise ratio (PSNR) between the computed focus function and the fitted model. This R-measure is compatible with any model of focus function. In addition, it efficiently integrates into the SFF pipeline by exploiting the focus information of the images (see Fig. 3), which leads to a reduced computational cost.

In the next section, a methodology for detecting and discarding the pixels where the depth estimation is unreliable based on the reliability measure introduced above is presented. In section 4, the suitability of the proposed measures for assessing the reliability of SFF is analyzed experimentally.

3.3. Depth-map carving

Either R1 or R2 can be applied to the input image sequence in order to assign a reliability value to each pixel (i, j). In particular, the values of the R-measure for all pixels can be interpreted as a gray-scale image in which each gray level is associated with the reliability of the depth estimation for the corresponding pixel: the depth of those pixels with a high R-measure are more likely to be estimated correctly. As a result, the depth-map can be carved by removing the pixels whose depth estimation is more likely to be inaccurate. In particular, all pixels whose reliability is below a predefined threshold, α, should be discarded.

The reliability threshold must be found experimentally as a result of a training process. For a given training set, the depth-map carving can be thought of as a two-class classification task where the classes correspond to those pixels that should be discarded from the depth-map and those that should be kept. Thus, the threshold value α is selected so that the highest classification rate in the training set is obtained in order to maximize the classification accuracy. As in typical classification tasks, accuracy corresponds to the percentage of correctly classified pixels with respect to the total number of pixels in each image.

4. Experiments

The experiments presented in this work have been conducted in three stages. In the first stage, the working principle of the proposed reliability measures is illustrated. In the second stage, the proposed R-measures have been applied in order to predict the quality of the obtained depth-maps. The objective of this stage is to demonstrate that the proposed reliability measures can be used for accurately detecting low-reliability regions in a depth-map in order to discard them. In the third stage, the proposed reliability measure is compared with different alternatives for carving the depth-map and removing inaccurate pixels.

Both synthetic and real sequences have been used. All simulated scenes consist of 25 images of 640×640 pixels. Defocus has been simulated for a 3.3 mm focal length camera focusing between 50 mm and 200 mm. Different surface shapes located between 100 mm and 150 mm away were mapped with different textures for each scene. Different textures, shapes and noise levels have been selected so that a variety of features and imaging conditions were considered.

The real focus sequences have been captured with a Sony SNC-RZ50P camera. Every sequence consists of 50 images of 640×480 pixels acquired at different focus ranges within 2m and 14m away from the camera, and with different targets. A total of 12 synthetic and 12 real sequences have been used in the experiments[1].

4.1. Reliability measure

According to section 3, R1 and R2 predict the reliability of the estimated pixel depth. A third reliability measure, R3, is defined by using the model for the focus function proposed in [15]. The different R-measures are summarized in table 1. In order to illustrate the working principle of the proposed R-measures and assess their effectiveness, the following experiment is conducted.

---

[1]Images and further details can be found at `http://www.sayonics.com/research/reliability_measure.html`.

Table 1: Reliability measures

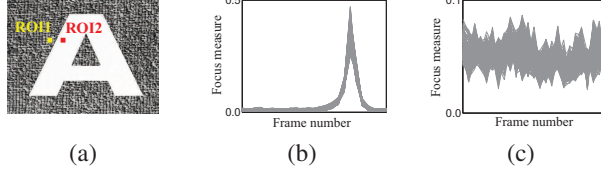|	Definition ($R_{i,j}$)	Model ($G_{i,j}$)
R1	$1 - \max\left(|F(x|G) - F(x|f)|\right)$	$A\exp\left((x-\mu)^2/(2\sigma^2)\right)$
R2	$20\log\left(f_{\max}/e_{i,j}\right)$	$A\exp\left((x-\mu)^2/(2\sigma^2)\right)$
R3	$20\log\left(f_{\max}/e_{i,j}\right)$	$A/(B + (C-k)^2)$[15]

(a)   (b)   (c)

Fig. 4: Effect of texture on the focus function. (a) Synthetic scene. The red and blue squares correspond to the first and second selected region, respectively. (b) Focus functions of the first region. (c) Focus functions of the second region.

Fig. 4(a) shows the all-in-focus image of a synthetic sequence. The labeled squares correspond to two manually selected regions of interest of 16×16 pixels, namely ROI1 and ROI2. Fig. 4(b) and 4(c) show the plots of all the focus functions corresponding to ROI1 and ROI2, respectively. Notice that each region of interest contains 256 focus functions: one for each pixel. It is evident that the focus functions corresponding to pixels from the low-textured region (ROI2) yield profiles that are unlikely to be modeled accurately. Therefore, the depth estimation using SFF is unreliable. In contrast, pixels from high-textured regions (ROI1) yield focus functions that can readily be modeled more accurately.

This behavior is illustrated in more detail in Fig. 5. This figure plots the measured focus functions and the corresponding focus models for the textures of Fig. 2(a). In this figure, it is clear that for the focus function corresponding to the pixels of a high-textured region, the theoretical model (red curves) adjusts well to the empirical data (black curves). In contrast, for the plots corresponding to pixels from the low-textured region, the reliability measures yield lower values since the real data depart from the fitted model.

The previous examples correspond to a rather simple case with clearly distinguishable regions that illustrate the working principle of the R-measures. In the next sections, the R-measures are extensively tested and compared with different alternatives.

4.2. Depth-map filtering

In this section, the reliability measure is applied in order to determine the regions where SFF is unlikely to perform correctly. For each reconstructed sequence, the aim is to generate a binary segmentation mask that removes the pixels whose depth error is above a predefined tolerance, say e_T.

In synthetic sequences, the ground truth is accurately known. Therefore, the error in the depth-maps can be readily computed. In this case, a reference segmentation mask, $M_{ref}(x, y)$, is gen-

erated for each scene as:

$$M_{ref}(x, y) = \begin{cases} 1 & \text{if } |Z(x,y) - GT(x,y)| > e_T \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

where $Z(x, y)$ is the depth-map, $GT(x, y)$ the corresponding ground-truth and e_T is the maximum error allowed for depth estimation (any pixel with a depth error higher than e_T is considered to be erroneous and discarded). For illustration, Fig. 6 shows the generation of the reference mask for the scene of Fig. 4. In this figure, pixels with erroneous depth estimate (error greater than e_T with respect to the ground truth) are marked with 1 (white) in the binary reference mask.

For complex real scenes, it is difficult to determine the ground truth accurately. In order to overcome this problem, the reference mask is manually generated by pre-computing the depth-map and marking those pixels whose estimated depth is incorrect. For a fair analysis of the results, a single reference mask was created for each scene and used in all the experiments.

Once the reference mask is generated for every scene, a segmentation mask, $M(x, y)$, is obtained by applying a threshold to the reliability measure. In the ideal case, the segmentation masks should be equal to the reference mask so that all pixels whose depth estimation is incorrect are removed from the depth-map.

In order to generate the segmentation, 3-fold cross-validation has been applied in section 3.3. The filtering quality is assessed by means of the accuracy, Acc:

$$Acc = 100 \sum_{(x,y)} \frac{\alpha(x,y)}{N}, \quad (15)$$

where the numerator corresponds to the coincident pixels between the actual segmentation mask and the reference one, and N is the total number of image pixels:

$$\alpha(x, y) = \begin{cases} 1 & \text{if } M_{ref}(x,y) = M(x,y) \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

In addition to (15), the precision (P) and recall (R) have been used for comparison:

$$P = \frac{tp}{tp + fp} \quad (17)$$

$$R = \frac{tp}{tp + fn}, \quad (18)$$

where tp and fp are the number of true positives and false positives, respectively and fn is the number of false negatives of M with respect to M_{ref}, with an error threshold of $e_T = 5\%$ in (14).

Table 2 shows the performance of the different reliability measures on synthetic scenes in terms accuracy, Acc, precision, P, recall, R, and the average area under the ROC curve, A (see Fig. 7). For all R-measures, the classification accuracy is above 85%. For illustration purposes, Fig. 8 shows the segmentation masks obtained by using the different reliability measures for the same scene of figures 4 and 6. In this figure, the red regions correspond to pixels that have incorrectly been classified.
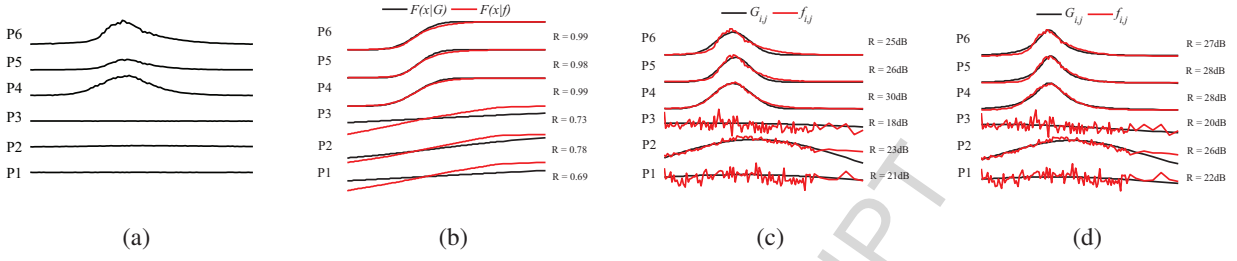
6

Fig. 5: Computation of reliability measures. (a) Focus functions corresponding to Fig. 2. (b) Fit using the Kolmogorov-Smirnov statistic. (c) Fit using the Gaussian model for the focus profile. (d) Fit using the Lorentzian-Cauchy model for the focus profile. For display purposes, the focus functions have been normalized between 0 and 1 in (b)-(c).
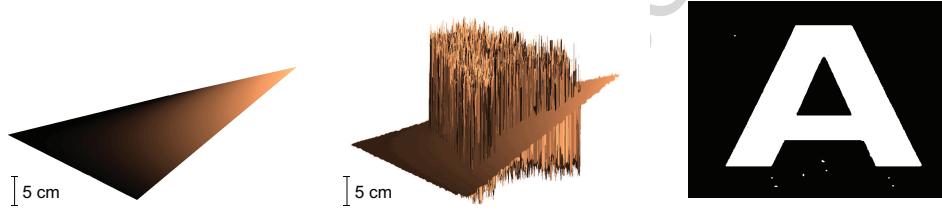

Fig. 6: Binary reference mask. From left to right: ground truth, depth-map and binary reference mask.

Table 2: Mean performance of different reliability measures. Gaussian focus profile (R1 and R2) and Lorentzian-Cauchy focus profile (R3).

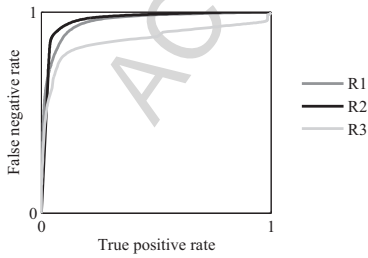| R-measure	Acc (%)	P(%)	R(%)	A
$R1$	86.2	65.3	91.4	0.9541
$R2$	89.5	70.0	91.8	0.9709
$R3$	85.3	60.2	85.2	0.8642

Fig. 7: Average ROC curves for 3-fold cross validation.

Fig. 8: Segmentation masks obtained from different reliability measures. From left to right: masks corresponding to R1, R2 and R3.

4.3. Performance comparison

The concept of reliability proposed in this work is aimed at providing a confidence value for each pixel of a given scene that measures the likelihood of obtaining an accurate depth estimate using SFF. In particular, the R2-measure has the advantage of being directly derived from the SFF algorithm by splitting the focus function into two components: an error signal and an ideal focus function model. As a result, the proposed reliability measure can be efficiently integrated into the traditional SFF algorithm and can readily be adapted to different focus function models. In order to assess the advantages and limitations of the proposed approach, this section compares the performance of depth-map filtering using the proposed reliability measure with three alternative filtering approaches. For the sake of brevity, only R2 has been used in the experiments. Based on the results presented in the previous section, R2 has been preferred over R1 and R3 due to its generality, simplicity and good performance.

Given the fact that texture is an important variable in the computation of depth maps using SFF, it is straightforward to exploit the texture information of the scene as a cue for computing an alternative reliability measure for SFF. However, in order to apply a texture segmentation approach, an all-in-focus image of the focus sequence is required. This implies that this alternative can only be applied as a post-processing by assuming that an accurate all-in-focus image of the scene can be obtained. The generation of all-in-focus images from a sequence of defocused images is an intensive research field. A discussion about focus fusion algorithms is out of the scope of this work. Different methodologies have been proposed based on image pyramids [30, 31], inverse filtering [14, 32, 33] and wavelet decomposition [34, 35]. For comparison purposes, the all-in-focus image

7

of each scene was computed using the *Helicon Focus* software [36] and has been fed into three alternative two-class texture classifiers.

It is important to remark that the generation of the all-in-focus image is not strictly necessary since the peak response of focus measure operators can be interpreted as a measure of the amount of texture. Notwithstanding, as remarked in [37], the neeed for generating an all-in-focus image is justified by the presence of false maxima in the focus functions corresponding to regions with low SNR. As a result, texture-based classifiers improve their performance if a high quality all-in-focus image is available.

A first texture classifier is simply obtained by applying the 24 Gabor filters described in [38] to the AIF image and then averaging the responses of all the filters for every pixel. Those filters are widely used for texture classification and segmentation. The average response of the filter bank is expected to yield high values in image regions with rich texture content and low values elsewhere. Thus, the average response is used to separate the image into two classes by simple applying a threshold. The threshold is selected by finding the value that yields the best classification rate in a training set. The second texture classifier is obtained by combining the responses of each individual Gabor filter using *Adaboost* [39, 40]. In this case, a *weak classifier* is generated by simply applying a threshold to the response of each filter. The threshold that yields the best classification rate is selected. Adaboost is then used to combine each weak classifier in order to obtain the best classification rate in the training set.[2]

In the experiments, a total of five filtering algorithms have been used in the comparisons: the Canny edge detector-based algorithm proposed in [22] (CAN), the mean response of the Gabor filters (GAB), the combination of Gabor filters using Adaboost (G+AD), the depth-map filtering-based algorithm proposed in [20] (DFIL) and the proposed reliability-based method (R2). In order to compare the effect of only the filtering stage of these approaches, all the depth-maps have been computed with the same SFF algorithm [7].

The different filtering algorithms require tuning their own parameters by means of a training process. For each filtering algorithm, the training process was carried out similarly as in the previous section for the synthetic sequences. In particular, for the real sequences, k-fold cross-validation has been applied in the training stage. The parameters of each filtering algorithm were adjusted in order to obtain the best classification rate in terms of accuracy. In particular, in the cross-validation process for R2, the segmentation threshold α varied between 12.3 and 15.3 dB in the real sequences.

Table 3 compares the mean performance of the different filtering methods. Fig. 9 and 10 show a frame of the focus sequence and the filtered depth-maps using the evaluated algorithms for three sequences from the synthetic test set and three sequences from the real test set, respectively.

---

[2]The Adaboost implementation used in this work can be found on line at http://www.mathworks.com/matlabcentral/fileexchange/27813.

Table 3: Mean performance of different filtering methods using 12 real sequences and 8 simulated sequences with 4-fold cross-validation. Rank of the algorithm in parenthesis.

(a) Real sequences

Method	Acc(%)	P(%)	R(%)
CAN	61.0 (5)	41.6 (4)	99.7 (1)
GAB	78.1 (3)	56.8 (3)	83.6 (4)
G+AD	86.7 (2)	67.4 (2)	88.4 (3)
DFIL	63.8 (4)	38.3 (5)	27.1 (5)
R2	88.4 (1)	68.1 (1)	93.4 (2)

(b) Simulated sequences

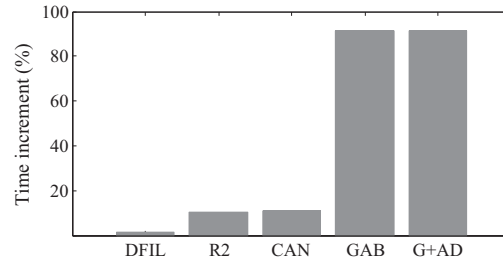| Method	Acc(%)	P(%)	R(%)
CAN	66.6 (4)	36.5 (4)	97.3 (1)
GAB	75.0 (3)	48.5 (3)	73.6 (4)
G+AD	87.1 (2)	72.6 (1)	75.1 (3)
DFIL	65.6 (5)	30.2 (5)	51.0 (5)
R2	89.5 (1)	70.0 (2)	91.8 (2)

Fig. 11: Computational cost of different filtering methods.

The Matlab implementation of the SFF routine runs in approximately 4.05s for a sequence of 25 images of 640×640 pixels on an Intel 2 Quad processor at 2.5GHz and 4GB of RAM. Fig. 11 summarizes the time increment (as a percentage of the duration of the original SFF routine) for different alternatives. The computation time of the all-in-focus image for texture-based methods (CAN, GAB and G+AD) has not been included in Fig. 11 since a third-party software has been used for its computation.

5. Discussion

The reliability measure, R1, presented in section 3.2 is aimed at predicting the confidence of SFF for estimating the depth of each pixel of an imaged scene. This R-measure is based on the Kolmogorov-Smirnov statistic and exploits the Gaussian model of the focus function. A more general R-measure, compatible with any model of the focus function, is proposed in (13). The experiments conducted in section 4.1 illustrate the working principle behind the concept of the R-measure. The results obtained in this section show that the R-measure effectively assigns high reliability values in those regions where SFF is likely to have a good performance without having neither the final depth-map nor the all-in-focus image.
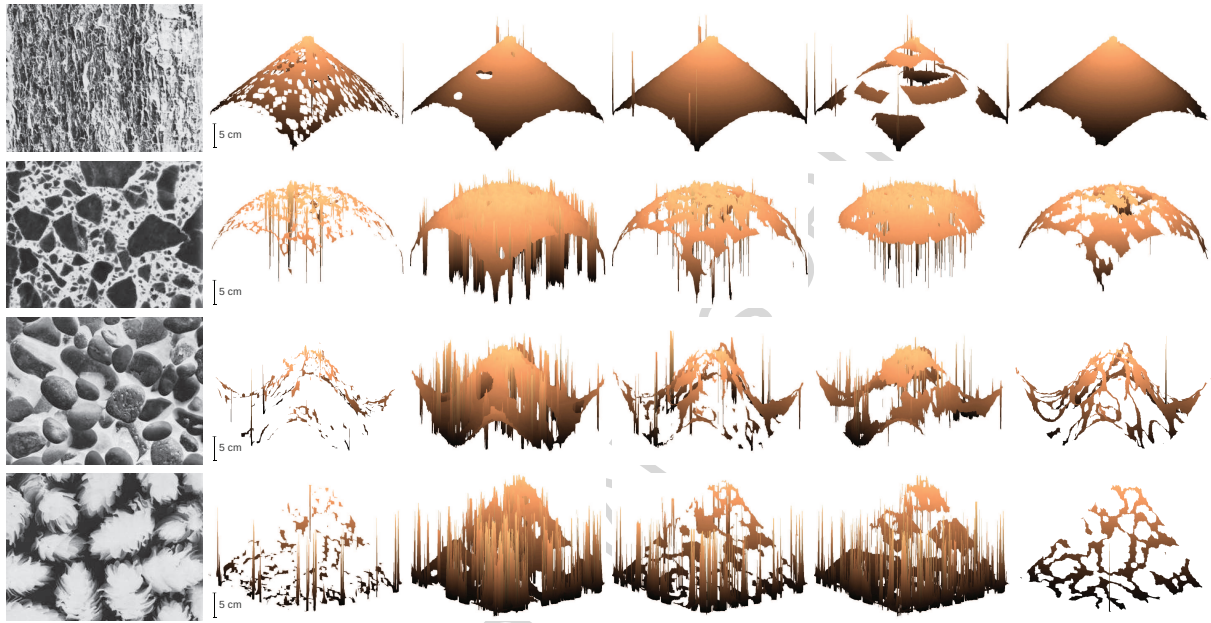
Fig. 9: Performance comparison with synthetic sequences. From left to right: All-in-focus image, and depth-maps filtered with CAN, GAB, G+AD, DFIL and REL, respectively.
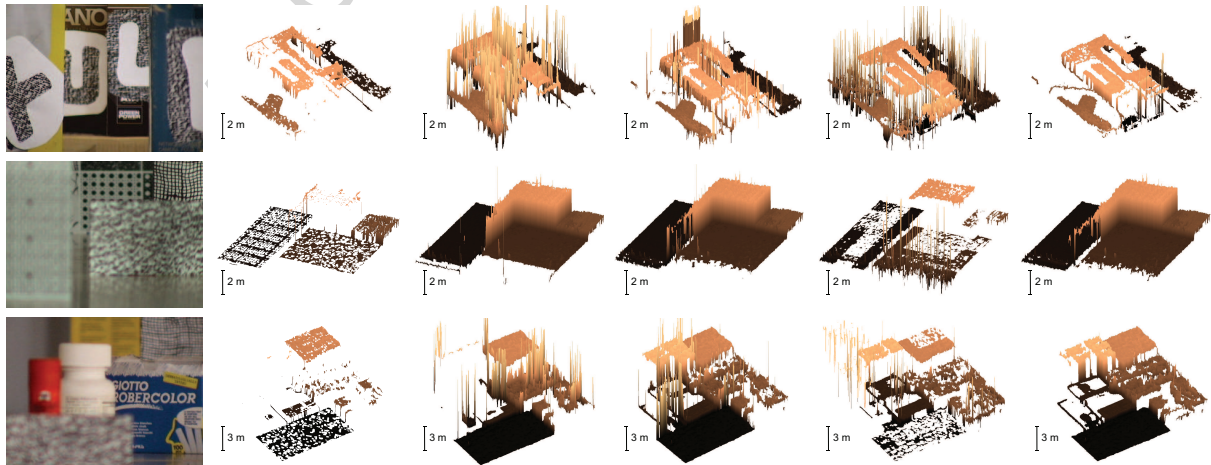

Fig. 10: Performance comparison with real sequences. From left to right: All-in-focus image, and depth-maps filtered with CAN, GAB, G+AD, DFIL and REL, respectively.

A straightforward application of the *R*-measure consists of using it for detecting pixels whose reliability is below a given threshold in order to discard them while preserving the useful information of the depth-map of complex scenes. The results presented in sections 4.2 and 4.3 show the advantages of the *R*-measure with respect to different alternatives. In table 3, the cascade classifier based on Adaboost showed the best performance for the synthetic scenes in terms of precision. Notwithstanding, the *R*-measure outperforms all the other methods in terms of accuracy in synthetic sequences and accuracy and precision in the real sequences. In addition, it ranks between the first and second place in all the quality measures for both the synthetic and real sequences. Some methods, such as CAN, provide a high recall at the cost of low accuracy and precision. This behavior is best illustrated in Fig. 9 and Fig. 10, in which some methods over-carve the depth-map, thus yielding a high recall at the cost of removing relevant information. Alternatively, some methods allow too much noise in the carved depth-map. In general, the proposed *R*-measure yields a good tradeoff between the different quality measures by removing erroneous pixels while preserving the relevant information of the depth-map.

Texture-based methods (e.g., CAN, GAB and G+AD) base their response on the texture information of the all-in-focus images with an important drawback: the all-in-focus image does not take into account the variations of the focus function along the *z*-axis (as a function of the in-focus position), which is an important factor in the performance of SFF. These variations are mainly due to CCD noise or optical effects such as the curvature field, image shift or artifacts. This could explain why they perform better only in the ideal case (in the synthetic sequences). In contrast, the proposed *R*-measure performs satisfactorily in both synthetic and real sequences.

Two of the main advantages of the proposed *R*-measure are its simplicity and efficient integration into SFF, which lowers the computational cost of SFF. Fig. 3 shows the *R*-measure within the SFF pipeline. According to Fig. 11, the computation of the *R*-measure yields an increase of approximately 10.6% in the computation time of the basic SFF stage. In addition, in the experiments with real scenes, the reliability threshold, α, had a little variation (between 12.3 and 15.3 dB). This is desirable since it suggests that the *R*-measure readily adapts to different imaging conditions and scenes without significant changes on its behavior.

Improving the robustness of SFF to different imaging factors and acquisition conditions is fundamental for enhancing the quality of the obtained depth-maps. Complementarily, the proposed *R*-measure is aimed at predicting the performance of the depth estimation in order to take advantage of the state-of-the-art SFF technique.

6. Conclusions

The concept of reliability for shape-from-focus has been introduced in this work. A reliability measure, *R-measure*, aimed at predicting the confidence of the depth-estimation using SFF

has been presented. The proposed *R*-measure efficiently exploits the information of the focus signals corresponding to each pixel of the imaged scene in order to compute a reliability value.

An application of the proposed *R*-measure for detecting and removing low-accuracy regions of the reconstructed depth-maps has been analyzed. The proposed approach is a step toward the application of SFF to complex scenes without previous knowledge or restrictions on the image content. In the literature, most results of applying SFF to real sequences have been limited to microscopic scenes, where an accurate control of the focus position and a shallow depth of field are more easily attainable. Alternatively, the proposed methodology provides a scheme for applying SFF to real macroscopic scenes by selecting the regions where SFF will yield accurate results. The proposed technique has been validated with both synthetic and real data and compared with different alternatives, showing a significantly higher performance.

Future work will focus on denoising and recovering low-reliability regions of the filtered depth-maps. The problem of recovering lost data in images [41, 42] and surfaces [43, 44] is referred to as *hole filling* or *surface filling*. Intensive research is being devoted to this topic in the computer vision community. In this way, the aim is to integrate the *R*-measure with existing filling and regularization techniques.

In this work, the *R*-measure has been applied successfully using the Gaussian model and the Lorentzian-Cauchy model for the focus profile. Notwithstanding, it can be readily extended to deal with different focus profiles. In this way, future research will be aimed at proposing new focus profile models for conventional cameras.

References

[1] Z. Jiang, D. Xu, M. Tan, H. Xie, An improved focus measure for MEMS assembly, in: Int. Conference on Mechatronics and Automation, Vol. 2, 2005, pp. 1118–1122.

[2] M. Watanabe, S. Nayar, Telecentric optics for constant magnification imaging, Tech. rep., Dept. of Computer Science, Columbia University (1995).

[3] S. K. Nayar, M. Watanabe, M. Noguchi, Real-time focus range sensor, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (1996) 1186–1198.

[4] H. Y. Lin, M. Subbarao, A vision system for fast 3d model reconstruction, in: Proc. IEEE Computer Society Conference on Computer Vision Pattern Recognition, Vol. 2, 2001, pp. 663 – 668. `doi:10.1109/CVPR.2001.991027`.

[5] K. Pradeep, A. Rajagopalan, Improving shape from focus using defocus cue, IEEE Trans. Image Process. 16 (7) (2007) 1920 –1925. `doi:10.1109/TIP.2007.899188`.

[6] K. Pradeep, A. Rajagopalan, Improving shape from focus using defocus information, in: proc. International Conference on Pattern Recognition, Vol. 1, 2006, pp. 731–734. `doi:10.1109/ICPR.2006.704`.

[7] S. K. Nayar, Shape from focus, Tech. Rep. CMU-RI-TR-89-27, Carnegie Mellon University, Pitsburg, PA (November 1989).

[8] S. K. Nayar, Y. Nakagawa, Shape from focus, IEEE Trans. Pattern Anal. Mach. Intell. 16 (8) (1994) 824–831. `doi:10.1109/34.308479`.

[9] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: proc. International Conference on Computer Vision, 1998, pp. 839 –846. `doi:10.1109/ICCV.1998.710815`.

[10] A. Buades, B. Coll, J.-M. Morel, A non-local algorithm for image denoising, in: proc. IEEE Computer Society Conference on Computer Vi-

10

sion and Pattern Recognition, Vol. 2, 2005, pp. 60 – 65 vol. 2. `doi:10.1109/CVPR.2005.38`.

[11] A. Buades, B. Coll, J. M. Morel, A review of image denosing algorithms, with a new ona, Multiscale Model. Simul. 4 (2005) 490–530.

[12] B. Jahne, Practical Handbook on image processing for scientific and thechnical applications, 2nd Edition, CRC Press, 2004.

[13] X. Hu, P. Mordohai, A quantitative evaluation of confidence measures for stereo vision, Pattern Analysis and Machine Intelligence, IEEE Transactions on 34 (11) (2012) 2121–2133. `doi:10.1109/TPAMI.2012.46`.

[14] M. Subbarao, T. Choi, Accurate recovery of three-dimensional shape from image focus, IEEE Trans. Pattern Anal. Mach. Intell. 17 (3) (1995) 266–274. `doi:10.1109/34.368191`.

[15] M. Muhammad, T.-S. Choi, Sampling for shape from focus in optical microscopy, IEEE Trans. Pattern Anal. Mach. Intell. 34 (3) (2012) 564 –573. `doi:10.1109/TPAMI.2011.144`.

[16] M. Asif, T. S. Choi, Shape from focus using multilayer feedforward neural networks, IEEE Trans. Image Process. 10 (11) (2001) 1670–1675. `doi:10.1109/83.967395`.

[17] M. B. Ahmad, T. S. Choi, Shape from focus using optimization technique, in: proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, Vol. 2, 2006, pp. II –II. `doi:10.1109/ICASSP.2006.1660387`.

[18] H. Sundaram, S. Nayar, Are textureless scenes recoverable?, in: proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, pp. 814–820.

[19] H. Shoji, K. Shirai, M. Ikehara, Shape from focus using color segmentation and bilateral filter, in: 4th - Digital Signal Processing Workshop, 12th - Signal Processing Education Workshop, 2006, pp. 566–571. `doi:10.1109/DSPWS.2006.265487`.

[20] M. Muhammad, H. Mutahira, A. Majid, T.-S. Choi, Recovering 3d shape of weak textured surfaces, in: Int. Conference on Computational Science and Its Applications, 2009, pp. 191 –197. `doi:10.1109/ICCSA.2009.25`.

[21] V. Gaganov, A. Ignatenko, Robust shape from focus via markov random fields, in: proc. Int. Conference on Computer Graphics and Vision, 2009, pp. 74–80.

[22] M. Muhammad, T.-S. Choi, An unorthodox approach towards shape from focus, in: IEEE Int. Conference on Image Processing, 2011, pp. 2965 –2968. `doi:10.1109/ICIP.2011.6116283`.

[23] N. N. K. Chern, P. A. Neow, M. H. Ang, Practical issues in pixel-based autofocusing for machine vision, in: proc. IEEE Int. Conference on Robotics and Automation, Vol. 3, 2001, pp. 2791–2796. `doi:10.1109/ROBOT.2001.933045`.

[24] R. Minhas, A. A. Mohammed, Q. J. Wu, Shape from focus using fast discrete curvelet transform, Pattern Recognit. 44 (4) (2011) 839 – 853. `doi:10.1016/j.patcog.2010.10.015`.
URL `http://www.sciencedirect.com/science/article/pii/S0031320310004991`

[25] S. Pertuz, D. Puig, M. A. Garcia, Analysis of focus measure operators for shape-from-focus, Pattern Recognition 46 (5) (2013) 1415–1432. `doi:10.1016/j.patcog.2012.11.011`.

[26] M. Subbarao, T. Choi, A. Nikzad, Focusing techniques, Journal of Optical Engineering 32 (1993) 2824–2836. `doi:10.1117/12.147706`.

[27] J. Dias, A. de Almeida, H. Araujo, Depth recovery using active focus in robotics, in: proc. IEEE/RSJ Int. Workshop on Intelligent Robots and Systems, Vol. 1, 1991, pp. 249–255. `doi:10.1109/IROS.1991.174458`.

[28] E. Wong, A new method for creating a depth map for camera auto focus using an all in focus picture and 2D scale space matching, in: proc. IEEE Int. Conference on Acoustic, Speech and Signal Processing, Vol. 3, 2006. `doi:10.1109/ICASSP.2006.1660871`.

[29] D. C. Tsai, H. H. Chen, Reciprocal focus profile, IEEE Trans. Image Process. 21 (2) (2012) 459 –468. `doi:10.1109/TIP.2011.2164417`.

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

[31] Z. Zhang, Flexible camera calibration by viewing a plane from unknown orientations, in: proc. IEEE International Conference on Computer Vision, Vol. 1, 1999, pp. 666 –673. `doi:10.1109/ICCV.1999.791289`.

[32] K. Kodama, H. Mo, A. Kubota, Simple and fast all-in-focus image reconstruction based on three-dimensional/two-dimensional transform and filtering, in: proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, 2007, pp. 769–772. `doi:10.1109/ICASSP.2007.366021`.

[33] F. Aguet, D. Van De Ville, M. Unser, Model-based 2.5-D deconvolution for extended depth of field in brightfield microscopy, IEEE Trans. Image Process. 17 (7) (2008) 1144–1153.

[34] B. Forster, D. Van De Ville, J. Berent, D. Sage, M. Unser, Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images, Microscopy Research and Technique 65 (1-2) (2004) 33–42.
URL `http://bigwww.epfl.ch/publications/forster0404.html`

[35] J. Tian, L. Chen, Multi-focus image fusion using wavelet-domain statistics, in: proc. IEEE Int. Conference on Image Processing, 2010, pp. 1205 –1208. `doi:10.1109/ICIP.2010.5651791`.

[36] Helicon Soft, Helicon focus, On line, available: http://www.heliconsoft.com/heliconfocus.html (Abr 2011).
URL `http://www.heliconsoft.com/heliconfocus.html`

[37] S. Pertuz, D. Puig, M. Garcia, A. Fusiello, Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images, IEEE Transactions on Image Processing 22 (3) (2013) 1242–1251. `doi:10.1109/TIP.2012.2231087`.

[38] B. Manjunath, W. Ma, Texture features for browsing and retrieval of image data, IEEE Trans. Pattern Anal. Mach. Intell. 18 (8) (1996) 837 –842. `doi:10.1109/34.531803`.

[39] Y. Freund, R. Schapire, A desicion-theoretic generalization of on-line learning and an application to boosting, in: P. Vitanyi (Ed.), Computational Learning Theory, Vol. 904 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 1995, pp. 23–37, 10.1007/3-540-59119-2-166.
URL `http://dl.acm.org/citation.cfm?id=646943.712093`

[40] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Ann. Stat. 28.

[41] L.-M. Po, S. Zhang, X. Xu, Y. Zhu, A new multidirectional extrapolation hole-filling method for depth-image-based rendering, in: proc. IEEE Int. Conference on Image Processing, 2011, pp. 2589 –2592. `doi:10.1109/ICIP.2011.6116194`.

[42] M. Bertalmio, L. Vese, G. Sapiro, S. Osher, Simultaneous structure and texture image inpainting, IEEE Trans. Image Process. 12 (8) (2003) 882 – 889. `doi:10.1109/TIP.2003.815261`.

[43] R. Sagawa, K. Ikeuchi, Hole filling of a 3D model by flipping signs of a signed distance field in adaptive resolution, IEEE Trans. Pattern Anal. Mach. Intell. 30 (4) (2008) 686 –699. `doi:10.1109/TPAMI.2007.70726`.

[44] C.-Y. Chen, K.-Y. Cheng, A sharpness-dependent filter for recovering sharp features in repaired 3D mesh models, IEEE Trans. Visualization and Computer Graphics 14 (1) (2008) 200 –212. `doi:10.1109/TVCG.2007.70625`.

Highlights

>The performance of focus measure depends on the optics and imaging conditions.
>The concept of reliability in focus measure is introduced.
>A method for computing the reliability of shape-from-focus is presented.
>The proposed reliability integrates efficiently to shape-from-focus.
>The proposed method is experimentally effective.