HAL
open science

Robust 3-D reconstruction of surfaces from image focus by local cross-sectional multivariate statistical analyses: Application to human ex vivo corneal endotheliums.

Mathieu Fernandes, Yann Gavet, Jean-Charles Pinoli

To cite this version:

Mathieu Fernandes, Yann Gavet, Jean-Charles Pinoli. Robust 3-D reconstruction of surfaces from image focus by local cross-sectional multivariate statistical analyses: Application to human ex vivo corneal endotheliums.. Medical Image Analysis, 2012, 16 (6), pp.1293-1306. 10.1016/j.media.2012.05.004 . hal-00733040

HAL Id: hal-00733040

https://hal.science/hal-00733040v1

Submitted on 20 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust 3-D reconstruction of surfaces from image focus by local cross-sectional multivariate statistical analyses: application to human *ex-vivo* corneal endotheliums

Mathieu Fernandes*, Yann Gavet, Jean-Charles Pinoli

École Nationale Supérieure des Mines de Saint-Étienne
CIS-LMPG CNRS
158 cours Fauriel, 42023 Saint-Étienne cedex 2, France

Abstract

The considered problem of 3-D reconstruction consists in computationally and passively recovering both topography and texture of a scene surface observed by optical sectioning with a limited depth-of-field imaging system (typically a conventional optical microscope). Throughout a sequence of registered 2-D images, the concepts of shape-from-focus and extended-depth-of-field respectively involve recovering both topography (depth map) and texture image of the surface by researching in-focus information. Toward that aim, traditional approaches generally follow a 2-D sectional way and thereby fail to deal with disturbed acquisitions, quite frequent in transmitted light observations and of interest in this paper. Such examples are the acquisitions from the medical issue addressed here: the 3-D reconstruction of human *ex-vivo* corneal endotheliums. These are mainly damaged by cellular fragments in the sample immersion medium or contrast reversals emphasized by both sample over-illumination and transmitted light imaging. To achieve with such noisy and disturbed acquisitions, a new focus analysis is introduced that originally adopts a 3-D strategy throughout the image sequence. This method exploits simultaneously all available cross-sectional cues to recover the observed surface that effectively strengthens the robustness. More precisely, it locally performs multivariate statistical analyses over cross-sectional spatial windows so as to find the sectional in-focus positions. Comparisons to state-of-the-art methods on both synthetic data and real acquisitions from the deal-with medical issue demonstrate the efficiency and the robustness of the proposed approach.

Keywords: Focus measurement, shape from focus, extended depth of field, robustness, eigenvalues decomposition, cornea endothelium.

1. Introduction

Observing through any optical imaging device with traditional lens and sensor plane system leads to information losses: (1) topographical information because of the projection on the sensor plane; (2) textural information "stained" by the focal range of the lens system. After the projection on the sensor plane, an observed surface scene whose profile covers more than the system depth-of-field (then describes as "thick") effectively appears in-focus and sharp on the acquired image only at parts lying within it. The remaining out-of-focus parts are blurred by the point spread function (PSF) of the system. In order to recover such a "thick" surface scene, one way can consist in gradually scanning it with the object focal plane of the imaging system, *i.e.* by acquiring a large sequence of images by optical sectioning (Agard, 1984). From such image sequence, this paper focuses on the computational and passive restoration of both topographical and textural information of the surface scene through the common complementary concepts of Shape-From-Focus (SFF) and Extended Depth-of-Field (EDF) respectively; the 3-D reconstruction being deduced by mapping the restored texture onto the estimated topography. In this work, we propose an effective theoretical improvement, subsequently applied to an original medical issue.

The SFF concept exploits the limited depth-of-field to infer the topography (or the shape) of the observed surface by axially (along the optical axis) maximizing a focus measurement throughout the image sequence (Nayar and Nakagawa, 1994), (Subbarao and Choi, 1995), (Malik and Choi, 2008), (Mahmood and Choi, 2008), (Mahmood et al., 2008), (Mahmood et al., 2009). Except that it is locally performed, the focus analysis is identical to autofocus algorithms that intend to automatically focus optical imaging systems (Brenner et al., 1976) (Krotkov, 1987), (Vollath, 1987), (Subbarao et al., 1993). Likewise, the EDF concept conversely tries to overcome the depth-of-field limitation by joining through a focus measurement the most focused information from the image sequence into a single image: the so-called texture image (Pieper and Korpel, 1983), (Sugimoto and Ichioka, 1985), (Valdecasas et al., 2001), sometimes through some form of multiresolution analysis (Forster et al., 2004), (Meneses et al., 2008). The latter are often introduced in the practical context of image fusion that consists in combining information from some (generally between 2 and

*Corresponding author.
Email addresses: mathieu.fernandes@gmail.com (Mathieu Fernandes), gavet@emse.fr (Yann Gavet), pinoli@emse.fr (Jean-Charles Pinoli)

5) multi-focus or multimodal images of the same scene into a single composite representation (Zhang and Blum, 1999). Recently, Aguet et al. (2008) proposed a model-based EDF method that therefore do not require any focus measurements. They estimate both textural and topographical information by globally fitting a convolutional model of defocus to the image sequence, as an iterative energy minimization problem. Despite this latter, the major part of SFF/EDF methods foremost rely upon an essential step of focus measurement to detect in-focus portions of the images. Toward that end, prominent cues are the high-spatial-frequency components of the surface texture, corresponding to sharpness information less damaged by the PSF. Nevertheless, this high-pass criterion can be corrupted by noise introduced by imaging equipments during acquisition. Indeed, most noises also result in high-spatial-frequency components of an image (Gonzalez and Woods, 2008) and thus provide false information of sharpness that misleads and affects the reconstruction process. From a theoretical point of view, this work is more particularly interested in developing a novel focus measurement offering a large noise robustness to make noisy reconstructions possible. To that aim, we originally adopt a 3-D strategy within the image sequence and perform local cross-sectional multivariate statistical analyses. Moreover, this noisy issue is even more frequent with reconstructions of transparent surfaces through transmitted light observations, *e.g.* common in medical and biological issues using conventional optical microscopy. To deal with such cases, SFF/EDF methods necessarily require that only the transparent surface appears in-focus throughout the image sequence, but it usually arises disturbed in there. Such examples are the acquisitions from the practical medical issue presented in this paper: the 3-D reconstruction of human *ex-vivo* corneal endotheliums observed by conventional optical microscopy.

The cornea is the transparent tissue covering the front of the eye. It assumes a role of protection of the eye and, together with the crystalline lens, transmits and focuses incident light onto the retina. It is arranged in several basic layers: such as, from outside to inside, the epithelium, the stroma and the endothelium. The latter thin innermost layer is constituted of essential cells tiled in a monolayer and hexagonal mosaic. Indeed, they pump water from the stroma, thus keeping the cornea clear. Nevertheless, they stretch to compensate for dead cells instead of regeneration, no longer maintaining a proper pumping action leading to a subsequent loss of stromal transparency. Thereby, a corneal transplantation may lastly be operated, before which a careful quality control of the corneal graft endothelium must be carried out. A high homogeneous density and a regular morphometry of the cells reflect the good quality of a cornea graft endothelium. Nowadays, the endothelium is inspected by observing the graft from the backside using conventional optical transmitted light microscopy and the cell density is estimated in 2-D by performing either a manual cell counting (Thuret et al., 2003) or a (semi)-automated image analysis (Gain et al., 2002), (Gavet and Pinoli, 2008). However, corneal grafts exhibit a natural convex shape and, most notably, posterior folds that necessarily appear during storage in a specific organ culture medium (Pels and Schuchard, 1983). Through the limited depth-of-field of the microscope, the endothelial mosaics are only examined in 2-D over restricted sharp regions that results in biased estimations of endothelial cell density (Thuret et al., 2003), (Thuret et al., 2004). From a practical point of view, this paper concerns the 3-D reconstruction of endothelial surfaces of human cornea grafts through the SFF/EDF concepts, in order to improve their quality control. As previously mentioned, the difficulty is that such transparent surfaces appear very disturbed throughout the acquired image sequence, in our case by some cellular fragments or by intense contrast reversals.

The paper is organized as follows. First, some image formation details and requirements are briefly highlighted (Section 2). Second, the SFF/EDF concepts are further explained and the main existing focus measurements are described (Section 3). Third, behind a brief review of statistical preliminaries (Section 4), our new 3-D EIG SFF/EDF approach is introduced (Section 5). Finally, the proposed 3-D EIG method is illustrated and validated on both synthetic data (Section 6) and real acquisitions from the just addressed medical issue (Section 7).

2. Image formation details and requirements

A simplified optical imaging system consisting of a convex lens objective (also acting as a diaphragm) and a sensor plane is depicted in Fig. 1. Let $(O, \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}, \vec{\boldsymbol{z}})$ denote a Cartesian coordinate system: O is the optical center and the z-axis is along the optical axis. Consider a thick scene surface, either opaque and observed in reflected light or transparent and observed in transmitted light, viewed as a 3-D surface distribution $\boldsymbol{\mathcal{S}}$ of radiance[1] in the direction of O. A surface point (x, y, z) is perspectively projected by the objective on the sensor plane at z_i through the transformation: $(x', y') = (-x.z_i/z, -y.z_i/z)$, which will be disregarded in the ensuing equations about image formation. This surface $\boldsymbol{\mathcal{S}}$ can be sectioned into a stack of successive sections $\boldsymbol{s}_z$ along the z-axis: $\boldsymbol{\mathcal{S}}(x, y, z) = \boldsymbol{s}_z(x, y)$.

Imaging with this system effectively presents a common characteristic: its limited depth-of-field d around its so-called object focal plane at z_o, becoming more restricted the larger the numerical aperture NA (and therefore the magnification γ) of the objective is[2]. Indeed, only surface sections $\boldsymbol{s}_z$ that lie within the depth-of-field appear acceptably sharp on the acquired image, as their light rays being sufficiently focused on the sensor plane. As for light rays arising from outside surface sections at $\delta z = z - z_o$, they distribute their energies over the sensor plane, causing the blur on the acquired image (Horn, 2001), (Born and Wolf, 1991). Contrary to EDF concept, the SFF one exploits this limitation and even rather requires a very restricted depth-of-field so as to perform well and to become accurate. By regarding the light illumination as incoherent, such

[1] In the context of radiometry, the term radiance refers to energy emitted along a certain direction, per solid angle, and per foreshortened area (Horn, 2001), (Born and Wolf, 1991).

[2] A Gaussian convex lens of focal length f theoretically focuses on a fixed image plane at z_i at the magnification $\gamma = z_i/z_o \sim$ NA only the light rays arising from a single object plane at z_o, the so-called object focal plane, obeying the Snell's formula: $1/z_i - 1/z_o = 1/f$ for the same medium refractive indexes in both front and back of the lens (Mahajan, 1998).

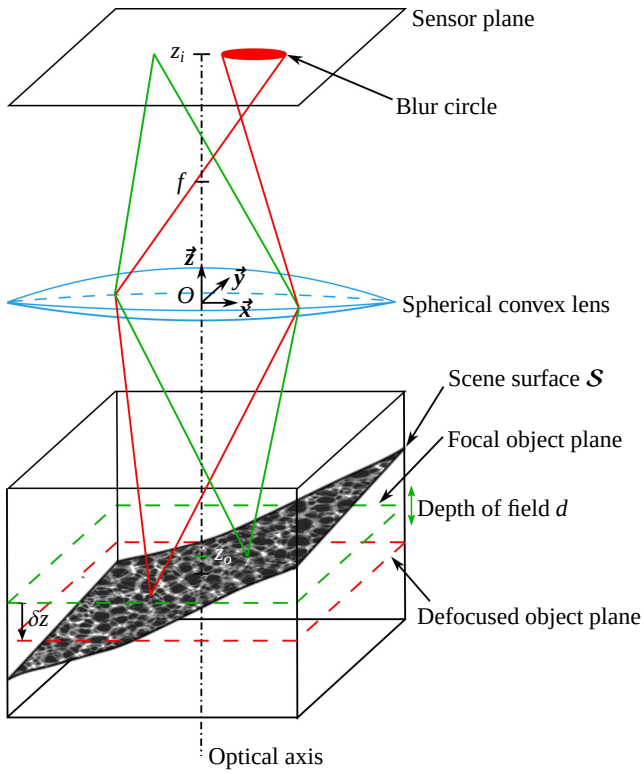


Figure 1: Illustration of the basic 2-D image formation geometry. The green light rays radiated by an in-focus point of $\mathcal{S}$ are well refracted onto the sensor plane contrary to the red light rays arising from an out-of-focus point, which converge forward and whose energies are distributed over the blur circle patch.

an aberration-free imaging system can be modelled as a 2-D shift-invariant linear system, *i.e.* in terms of convolutions with its impulsive response: the point spread function PSF (Born and Wolf, 1991), (Horn, 2001). The irradiance[3] (related to radiance, itself related to intensity) image acquired with the object focal plane set to z_o, denoted $\boldsymbol{\imath}|_{z_o}$, is thereby the contribution of each successive object section $\boldsymbol{s}_z$ convolved with its corresponding 2-D sectional PSF $\boldsymbol{h}_{\delta z}$ (Agard, 1984):

$$\boldsymbol{\imath}|_{z_o}(x,y) = \int_{-\infty}^{\infty} \boldsymbol{s}_z(x,y) * \boldsymbol{h}_{\delta z}(x,y)\, dz, \qquad (1)$$

where $\boldsymbol{h}$ denotes the 3-D PSF of the imaging system to a punctual object at (x, y, z_o).

The final sequence (3-D image) of 2-D images, denoted $\mathcal{I}$, is acquired by optical sectioning, *i.e.* by varying the z-position of the object focal plane by constant steps Δz throughout the surface $\mathcal{S}$: $\mathcal{I}(x,y,z) = \boldsymbol{\imath}|_z(x,y)$ (Agard, 1984). Both SFF/EDF concepts necessarily require a sequence with image sections spatially registered, principally by considering magnification variations due to changes in focus setting through the perspective projection of most optical imaging system[4] (as in Fig. 1) (Willson and Shafer, 1991). These magnification changes can be corrected using optical approaches, such as zoom adjustments based on system calibration (Willson, 1994), or computational techniques, commonly referred to as image warping (Darrell and Wohn, 1988). Notice that acquiring the image sequence by displacing either the scene or the imaging system along the z-direction with respect to a fixed focus setting ensures at least a constant magnification γ for all successive object focal planes, but not for the out-of-focus object planes that always suffer different magnifications than the focal ones (Nayar and Nakagawa, 1994). Otherwise, an all-over constant magnification can be reached through orthographic projection of telecentric optics (Watanabe and Nayar, 1997). If that is the case, as well as aforementioned 2-D constraints, the imaging system can be considered as a 3-D shift-invariant linear system (Frieden, 1967), (Streibl, 1984) and the 3-D image $\mathcal{I}$ can be expressed by the following 3-D convolution (Agard, 1984):

$$\mathcal{I}(x,y,z) = \mathcal{S}(x,y,z) * \boldsymbol{h}(x,y,z), \qquad (2)$$

illustrated in Fig. 2.

Before tackling the reconstruction process, let us briefly describe the effects produced by the convolution with the PSF in Eq. 1 and Eq. 2. The PSF results from the contribution of many blur factors, such as the defocusing, the optical diffraction and aberrations and the sampling, principally (Mahajan, 1998), (Mahajan, 2001). Many theoretical models of PSF have been proposed, whose accuracies depend on considered factors and used approximations[5] (Born and Wolf, 1991), (Castleman, 1996), (Gibson and Lanni, 1989). Introduced by Pentland (1987), a 2-D Gaussian function is often suggested as a PSF model with a standard deviation function of the distance of defocus δz. In summary, the PSF always behaves as a low-pass filter, whose cut-off spatial frequency falls when the distance of defocus δz raises.

3. Previous work in shape-from-focus (SFF) / extended-depth-of-field (EDF)

Traditionally, SFF/EDF methods work by means of an essential step: a local focus measurement that consists in measuring the degree of focus of each voxel of the original image sequence $\mathcal{I}$. The topography, commonly referred to as depth map, and the texture image of the observed surface $\mathcal{S}$ are then inferred, preferably by taking the discretization of the acquisition process into account. Otherwise, Aguet et al. (2008) recently suggested a model-based approach, thereby free from focus measurement, which jointly estimates both topography and texture as a global optimisation problem.

[3] In the context of radiometry, the term irradiance refers to energy incident per foreshortened area (a pixel of the sensor plane) (Horn, 2001), (Born and Wolf, 1991).

[4] Since the intersections of the so-called principal rays (the ones passing undeflected through the center of the lens O) with the sensor plane vary with the position of this latter, the image magnification changes with defocus.

[5] According to geometrical optics, a first-order approximation of the defocusing PSF consists in a homogeneous patch, the so-called blur circle in the case of a spherical diaphragm whose radius increases with the distance of defocus δz (Horn, 2001).

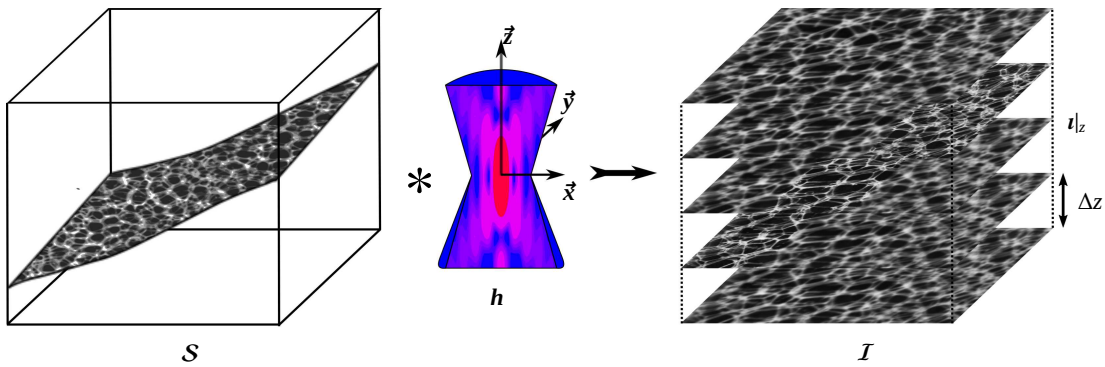


Figure 2: Illustration of the 3-D image formation model. The object surface $\boldsymbol{\mathcal{S}}$ is convolved with the 3-D PSF $\boldsymbol{h}$ of the imaging system to obtain a 3-D image $\boldsymbol{\mathcal{I}}$ as a sequence of 2-D images $\boldsymbol{\imath}|_z$ along the z-axis.

3.1. Focus-based approaches

3.1.1. Focus measurements

In view of the fact that the PSF acts as a low-pass filter, focus measurements thus try to locally emphasize and quantify high-spatial frequency components of the original image sequence $\boldsymbol{\mathcal{I}}$. Let $\boldsymbol{\mathcal{I}}(x,y,z)$ be defined on the spatial support $\mathbb{D} = \mathbb{D}_x \times \mathbb{D}_y \times \mathbb{D}_z \subset \mathbb{R}_+^3$ and valued into a positive real range $E \subset \mathbb{R}_+$ of intensity values. Applying a focus measurement function (FM) on it yields a 3-D focus degree measure $\boldsymbol{\mathcal{F}}(x,y,z)$ as follows:

$$\begin{aligned} \boldsymbol{\mathcal{F}} : \quad & \mathbb{D} \to \mathbb{R}_+ \\ & (x,y,z) \mapsto \mathrm{FM}(\boldsymbol{\mathcal{I}}(x,y,z)) \,, \end{aligned} \tag{3}$$

wherein the profile at location (x,y) along the z-direction is designated as $\boldsymbol{\mathcal{F}}|_{x,y} : \mathbb{D}_z \to \mathbb{R}_+$. Focus measurements can be classified according to the dimensionality of the adopted strategy to do that.

1-D approaches. From the early 1980s, some methods using maximum or minimum selection rules throughout single-voxel stacks along the z-direction of the image sequence $\boldsymbol{\mathcal{I}}$ are first proposed (Pieper and Korpel, 1983), (Sugimoto and Ichioka, 1985), but yielding non-robust approaches.

2-D approaches. For the last 40 years, a lot of more reliable focus measurements independently acting (in 2-D) on each image section of the sequence $\boldsymbol{\mathcal{I}}$ also arose. Classically, 2-D focus measurements work locally over sectional windows in two successive steps, aiming to emphasize and quantify focus cues respectively. The first step generally uses either high-pass filtering (Laplacian, Brenner or Tenenbaum gradients...) (Brenner et al., 1976), (Krotkov, 1987), (Subbarao et al., 1993), (Nayar and Nakagawa, 1994) or statistical tools (variance, autocorrelation or sum of eigenvalues...) (Sugimoto and Ichioka, 1985), (Vollath, 1987), (Wee and Paramesran, 2007), sometimes in various frequency domains (*e.g.* through discrete Fourier transform) (Malik and Choi, 2008) or with a different image processing framework (Fernandes et al., 2011a). To gain robustness, the second step simply consists in an energy measurement (commonly the sum over the considered windows of the absolute values) of the results of the first one. Otherwise, some 2-D methods operate through some forms of multiresolution decomposition analysis (*e.g.* Laplacian, gradient, ratio-of-low-pass pyramids or wavelet transform...) in order to perform high-pass filtering at different resolution level (Burt and Adelson, 1983), (Toet, 1989), (Forster et al., 2004). They work similarly except for the restoration of the texture image $\boldsymbol{\mathcal{T}}$ that is done by a reverse procedure after the fusion of most salient decomposition parts (Zhang and Blum, 1999). By independently working on each individual image section of the sequence $\boldsymbol{\mathcal{I}}$, 2-D approaches are inevitably misled by a rather isolated sectional noisy/disturbance data that appears sharpest, in theory more than the following 3-D approaches.

3-D approaches. Recently, Mahmood and Choi (2008) have introduced a 3-D focus measurement, *i.e.* which takes fully advantage of the three spatial dimensions of the original image sequence $\boldsymbol{\mathcal{I}}$. It is locally based on a principal component analysis within a stack of collected sectional neighborhoods along the z-direction. Consequently, it simultaneously exploits all focus cues along the axial (or cross-sectional) z-direction in order to estimate sectional degrees of focus. Contrary to 1-D/2-D ones, this novel 3-D strategy would enable to improve the robustness. However, we will see that it actually suffers from a severe loss of sensitivity. Indeed, it finally uses the largest principal component to discriminate in-focus information that represents the global content of the studied data. Hence, the authors combine it with various previous transforms, such as discrete wavelet (Mahmood et al., 2009) or cosine (Mahmood et al., 2008) transforms.

3.1.2. Topography and texture inferring

Topographical information. The z-coordinates of the voxels that exhibit the largest degrees of focus (referred to as depths) infer the topography (depth map) $\boldsymbol{\mathcal{D}}$ of the observed surface as follows:

$$\begin{aligned} \boldsymbol{\mathcal{D}} : \mathbb{D}_x \times \mathbb{D}_y \to \mathbb{D}_z \\ (x,y) \mapsto \operatorname*{argmax}_{z \in \mathbb{D}_z} \boldsymbol{\mathcal{F}}|_{x,y}(z) \,. \end{aligned} \tag{4}$$

Because of the significant thickness d of the depth-of-field, the recovered topography $\boldsymbol{\mathcal{D}}$ shows inherent "staircase" effects that must be compensated for by a process of interpolation embedded in this basic procedure of reconstruction. The traditional

one consists in fitting a Gaussian distribution, whose mean finally constitutes the interpolated depth value, to the three degrees of focus lying on the largest mode (Nayar and Nakagawa, 1994). Similarly, a quadratic (or even more) polynomial model can be fitted, sometimes regarding more than three degrees of focus (Subbarao and Choi, 1995). A subsequent approach (based on approximation of the Focussed Image Surface (FIS)) locally tries to refine the initial recovered topography $\mathcal{D}$ by optimizing both position and orientation of 2-D planar (then curved) windows to maximize the covered degrees of focus throughout the 3-D measure $\mathcal{F}$ (Subbarao and Choi, 1995), (Yun and Choi, 1999), (Asif and Choi, 2001), (Ahmad and Choi, 2005). Finally, the topography is often smoothed through average or median filters (Mahmood and Choi, 2008), (Mahmood et al., 2008). Interpolation techniques lying beyond the scope of this paper, only the traditional one will be used herein, finalised by a spline smoothing.

Textural information. Throughout the image sequence $\mathcal{I}(x, y, z)$, the texture image $\mathcal{T}$ of the observed surface is restored by joining the intensity voxels with the largest degrees of focus:

$$\begin{aligned} \mathcal{T} : \mathbb{D}_x \times \mathbb{D}_y &\to E \\ (x, y) &\mapsto \mathcal{I}(x, y, \operatorname*{argmax}_{z \in \mathbb{D}_z} \mathcal{F}|_{x,y}(z)) \,. \end{aligned} \tag{5}$$

When the optical sectioning step is larger than the depth-of-field ($\Delta z > d$), some regions of the observed surface may never appear in-focus throughout the image sequence and therefore on the restored texture image. Pradeep and Rajagopalan (2007) then proposed to perform a non-stationary Wiener filter to locally deconvolve the texture image $\mathcal{T}$. Note that no deconvolution process will be used herein.

Given the local and sectional action of the major part of the focus measurements, a certain depth regularity of the observed surface is implicitly assumed. On the one hand, the considered window has to be as small as possible to guarantee an approximate constant depth within itself and therefore to avoid too much smoothing the restoration process around sharp depth slopes or even depth discontinuities (Malik and Choi, 2007). On the other hand, it has to be sufficiently large to always capture focus cues within wide homogeneous textural contents and to average out noise.

3.2. Model-based approach

Lately, Aguet et al. (2008) suggested an approach (called 2.5-D deconvolution) based on the image formation model described by Eq. 2. The problem is stated as a least-squares minimization where the texture $\mathcal{T}$ and the depth map $\mathcal{D}$ are alternatively updated throughout $\mathcal{I}$. The latter is moreover estimated in a coarse-to-fine framework that imposes a certain regularity and robustness to the update. The PSF in Eq. 2 is modelled by a Gaussian function, whose standard deviation must be adjusted to produce satisfying results despite a relative insensitivity to this parameterization. During the optimization process, the current depth map $\mathcal{D}$ and texture image $\mathcal{T}$ are intrinsically interpolated and deconvolved, respectively.

This work aims at creating a novel 3-D focus measurement offering a large robustness to noise (unlike 2-D ones) while preserving a sufficient sensitivity to focus cues (unlike 3-D existing ones) in order to well operate through noisy/disturbed acquisitions, typically like the ones analysed in the dealt-with medical issue. Within the numerous SFF/EDF methods proposed in the specialized literature, some traditional and recent ones listed in Tab. 1 have been finally retained for the comparisons with our suggested 3-D approach.

4. Mathematical preliminaries

This section briefly reviews the basic theory of multivariate statistical analysis (Strang, 1980), (Fukunaga, 1990), (Jolliffe, 2002) used in the ensuing proposed SFF/EDF approach.

4.1. Multivariate data matrix

Let $\boldsymbol{X} \in \mathbb{R}^{m \times n}$ represents a $m \times n$ matrix of real-valued data, whose rows and columns are respectively denoted $\boldsymbol{X} = {}^t(\boldsymbol{r}_i)_{i \in [\![1,m]\!]}$ and $\boldsymbol{X} = (\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$, where t stands for the transpose operation. Consider a m-dimensional Hilbert space $\mathbb{R}^m$ whose the inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ defined as:

$$(\boldsymbol{c}_{j_1}, \boldsymbol{c}_{j_2}) \mapsto \langle \boldsymbol{c}_{j_1}, \boldsymbol{c}_{j_2} \rangle = \sum_{i=1}^{m} c_{ij_1} c_{ij_2}, \tag{6}$$

induces its norm $\| \cdot \|$, such that $\|\boldsymbol{c}\|^2 = \langle \boldsymbol{c}, \boldsymbol{c} \rangle$. The canonical basis of this Hilbert space $\mathbb{R}^m$ is denoted: $\boldsymbol{E} = (\boldsymbol{e}_j)_{j \in [\![1,m]\!]} = (\delta_{ij})_{(i,j) \in [\![1,m]\!]^2}$, where the columns $(\boldsymbol{e}_j)_{j \in [\![1,m]\!]}$ are the canonical vectors such that the Kronecker delta is $\delta_{ij} = 1$ if $i = j$, and $\delta_{ij} = 0$ otherwise. In the same way, a n-dimensional Hilbert space $\mathbb{R}^n$ can be considered.

Suppose that each column $(\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$ of $\boldsymbol{X}$ constitutes a set of m variates. The inertia of these column vectors $(\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$, denoted I, is defined as:

$$I = \frac{1}{m} \sum_{j=1}^{n} \|(\boldsymbol{c}_j - \overline{\boldsymbol{c}}_j)_{j \in [\![1,n]\!]}\|^2, \tag{7}$$

where the operator $\overline{\boldsymbol{c}}$ stands for the mean of the variate vector $\boldsymbol{c}$. Remark that this real-valued amount is thus equal to the sum of the variances of the column vectors $(\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$ of $\boldsymbol{X}$. It measures their dispersion in the Hilbert space $\mathbb{R}^n$.

4.2. Covariance matrix

Suppose that each variate column vector $(\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$ of $\boldsymbol{X}$ is centered: $\boldsymbol{X} = (\boldsymbol{c}_j - \overline{\boldsymbol{c}}_j)_{j \in [\![1,n]\!]}$. The covariance matrix of the columns $(\boldsymbol{c}_j)_{j \in [\![1,n]\!]}$ of $\boldsymbol{X}$, *i.e.* the covariance matrix $\boldsymbol{C_X}$ of the variables of $(\boldsymbol{r}_i)_{i \in [\![1,m]\!]}$, is the following $n \times n$ real square symmetric matrix:

$$\boldsymbol{C_X} = \frac{1}{m} \, {}^t\boldsymbol{X}\boldsymbol{X} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{r}_i \, {}^t\boldsymbol{r}_i, \tag{8}$$

2-D VAR	Variance within a 2-D window (Sugimoto and Ichioka, 1985)
2-D TEN	Sum over a 2-D window of the squared responses of the horizontal and vertical Sobel masks(Krotkov, 1987)
2-D OPT	Sum over a 2-D window of the absolute real responses of an "optical" band-pass filter (Malik and Choi, 2008)
2.5-D DEC	Model-based 2.5-D deconvolution method (Aguet et al., 2008)
3-D DCT-PCA	Principal component analysis of coefficients from discrete cosine transforms of 2-D/3-D windows collected along the z-direction (Mahmood et al., 2008)

Table 1: Designations, details and references of some retained state-of-the-art SFF/EDF methods.

The ij-th element, proportional to the inner product $\langle \boldsymbol{c}_i, \boldsymbol{c}_j \rangle$, corresponds to the covariance between $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ if $i \neq j$ or the variance of $\boldsymbol{c}_i$ otherwise. The inertia I of the columns $(\boldsymbol{c}_j)_{j\in[\![1,n]\!]}$ is simply the trace of $\boldsymbol{C_X}$: $I = \mathrm{tr}(\boldsymbol{C_X})$.

4.3. Eigenvalues decomposition

The eigenvalues decomposition of the covariance matrix $\boldsymbol{C_X} \in \mathbb{R}^{n\times n}$ of the columns $(\boldsymbol{c}_j)_{j\in[\![1,n]\!]}$ of $\boldsymbol{X}$ satisfies the equation:

$$\boldsymbol{C_X G} = \boldsymbol{\Lambda G}, \tag{9}$$

with a diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ and a matrix $\boldsymbol{G} \in \mathbb{R}^{n\times n}$. The non-negative diagonal elements $(\lambda_{ij})_{i=j\in[\![1,n]\!]}$ of $\boldsymbol{\Lambda}$ are the eigenvalues of $\boldsymbol{C_X}$. The columns of $\boldsymbol{G} = (\boldsymbol{g}_j)_{j\in[\![1,n]\!]}$ are the eigenvectors, such that $\langle \boldsymbol{g}_{j_1}, \boldsymbol{g}_{j_2} \rangle = 1$ if $j_1 = j_2$, and $\langle \boldsymbol{g}_{j_1}, \boldsymbol{g}_{j_2} \rangle = 0$ otherwise. The eigenbasis $\boldsymbol{G} = (\boldsymbol{g}_j)_{j\in[\![1,n]\!]}$ forms a novel orthonormal basis for the rows $(\boldsymbol{r}_i)_{i\in[\![1,m]\!]}$ of $\boldsymbol{X}$. Each eigenvector $\boldsymbol{g}_k$ with $k \in [\![1,n]\!]$ is associated with a particular eigenvalue λ_{kk}, which reveals its considered amount among the inertia I of the columns $(\boldsymbol{c}_j)_{j\in[\![1,n]\!]}$, as well as the variance of the rows $(\boldsymbol{r}_i)_{i\in[\![1,m]\!]}$ projected onto itself:

$$\lambda_{kk} = \boldsymbol{g}_k \boldsymbol{C_X}\, {}^t\boldsymbol{g}_k = \frac{1}{m}\boldsymbol{g}_k \left(\sum_{i=1}^{m} \boldsymbol{r}_i\, {}^t\boldsymbol{r}_i \right) {}^t\boldsymbol{g}_k = \frac{1}{m}\sum_{i=1}^{m} \langle \boldsymbol{r}_i, \boldsymbol{g}_k \rangle^2. \tag{10}$$

Note that $I = \mathrm{tr}(\boldsymbol{C_X}) = \mathrm{tr}(\boldsymbol{\Lambda}) = \sum_{k=1}^{n} \lambda_{kk}$. By convention, the ordering of the eigenvectors is determined by high-to-low sorting of the associated eigenvalues, with the highest eigenvalue in the upper left index of the $\boldsymbol{\Lambda}$ matrix. Hence, the first eigenvector $\boldsymbol{g}_1$ accounts for a variance λ_1 as large as possible among the inertia I exhibited by $(\boldsymbol{c}_j)_{j\in[\![1,n]\!]}$. The succeeding eigenvectors then maximize their considered variances among the remaining inertia, in order and subject to the orthogonal condition.

5. 3-D multivariate statistical measurement of focus

Based on the mathematical theory just summarized, a new 3-D focus analysis designated as 3-D EIG is introduced. It consists in several successive steps locally performed:

1. creation of a multivariate data matrix;
2. diagonalization of the data matrix (eigenbasis);
3. dimension reduction of the eigenbasis;
4. focus measurement.

Each of them, among others, constitutes one of the following subsections.

5.1. Preliminary set definitions

Firstly, it is important to specify a couple of set definitions. Let $\mathfrak{P}(x,y)$ denotes a stack of single-voxels at (x,y) throughout the input 3-D image $\boldsymbol{\mathcal{I}}$ and along the z-direction:

$$\mathfrak{P}(x,y) = \left\{ (x', y', z'_p) \in \mathbb{D} \;/\; x' = x, y' = y, p \in [\![1,N]\!] \right\}, \tag{11}$$

whose cardinal number $\sharp\mathfrak{P}$ is equal to the number N of sections of $\boldsymbol{\mathcal{I}}$. Each voxel stack $\mathfrak{P}(x,y)$ materializes a particular orthographic projection ray of the image formation process.

Let $\mathfrak{B}_r(x,y,z)$ denotes the 2-D square-shaped neighborhood of the voxel (x,y,z) of $\boldsymbol{\mathcal{I}}$ defined by:

$$\begin{aligned} &\mathfrak{B}_r(x,y,z) = \\ &\{(x',y',z') \in \mathbb{D} \;/\; |x-x'| \vee |y-y'| \leqslant r, z' = z\}, \end{aligned} \tag{12}$$

with $\vee$ the maximum operator and r the size of the neighborhood. As a remark, it strictly lies on the 2-D image section $\boldsymbol{\iota}|_{\boldsymbol{z}}$ including the voxel (x,y,z) and is perpendicular to the projection ray $\mathfrak{P}(x,y)$. Several other combinations, like disk-shaped or rectangular-shaped, can be used for the selection of this 2-D sectional neighborhood.

5.2. Multivariate data matrix

In the following, let us consider one orthographic projection ray $\mathfrak{P}(x,y)$ at (x,y) within $\boldsymbol{\mathcal{I}}$. The $\sharp\mathfrak{B}_r$ voxels of the sectional neighborhood $\mathfrak{B}_r(x,y,z)$ of a particular voxel (x,y,z) of $\mathfrak{P}(x,y)$ can be firstly sorted in lexicographic order (see Fig. 3). Their respective intensities can then result in a column vector $\boldsymbol{b}_r(x,y,z)$:

$$\boldsymbol{b}_r(x,y,z) = {}^t\left(\boldsymbol{\mathcal{I}}(x',y',z')\right)_{(x',y',z')\in\mathfrak{B}_r(x,y,z)}. \tag{13}$$

The whole collection of resulting neighborhood vectors $\boldsymbol{b}_r(x,y,z)$ of the voxels composing the stack $\mathfrak{P}(x,y)$ forms the multivariate matrix $\boldsymbol{X}(x,y)$ (see Fig. 3) as follows:

$$\boldsymbol{X}(x,y) = \left(\boldsymbol{b}_r(x,y,z_p)\right)_{p\in[\![1,N]\!]}. \tag{14}$$

The i-th row $\boldsymbol{r}_i$ of $\boldsymbol{X}(x,y)$, referred to as the i-th cross-sectional response, is constituted by the same i-th component of all ordered neighborhood vector $\boldsymbol{b}_r(x,y,z)$. Alternatively, the j-th column $\boldsymbol{c}_j$ of $\boldsymbol{X}(x,y)$, referred to as the j-th sectional observation, forms a set of $\sharp\mathfrak{B}_r$ variates corresponding to all components of the particular ordered neighborhood vector $\boldsymbol{b}_r(x,y,z_j)$ (Fig. 3).

The inertia $I(x,y)$ of the sectional observations (*i.e.* the columns $(\boldsymbol{c}_j)_{j\in[\![1,N]\!]}$) of $\boldsymbol{X}(x,y)$ measures the dispersion of the

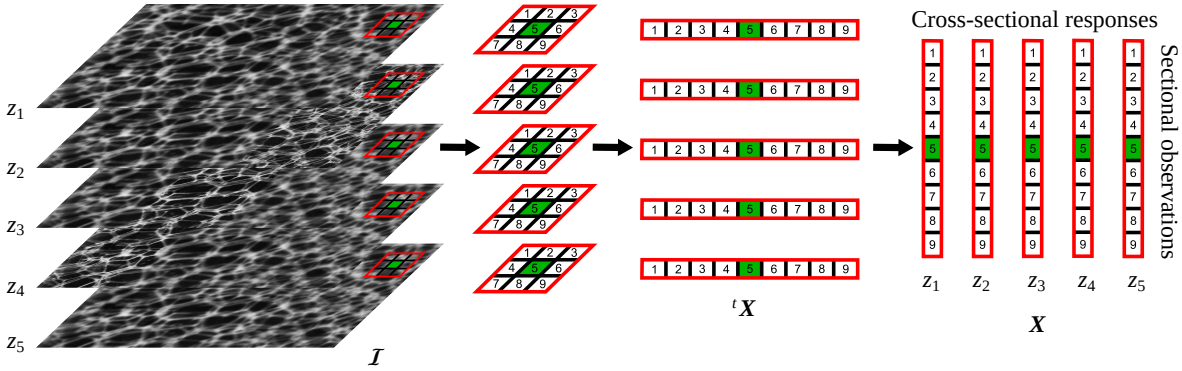


Figure 3: Illustration of the creation of the multivariate data matrix $\boldsymbol{X}$. In this simple example, the current voxel stack $\mathfrak{P}$ appears in green and every sectional neighborhoods $\mathfrak{B}_1$ of size 1 is bordered in red. Finally, the rows and the columns of the built matrix $\boldsymbol{X}$ are the cross-sectional responses and the sectional observations, respectively.

sectional observations of $\boldsymbol{X}(x,y)$ in the Hilbert space $\mathbb{R}^N$ of dimensions N, where $N = \sharp\mathfrak{P}$ is the number of sections of the sequence $\mathcal{I}$ (Eq. 7). Remark that this Hilbert space $\mathbb{R}^N$, in which the cross-sectional responses (*i.e.* the rows $(\boldsymbol{r}_i)_{i\in[\![1,N]\!]}$) of $\boldsymbol{X}(x,y)$ are defined, abstracts the image sequence $\mathcal{I}$. It can be naturally regarded as an "image sequence space", whose each dimension is referred to a different 2-D image section of the sequence $\mathcal{I}$ (see Fig. 3). Thus, each vector $\boldsymbol{e}_j$ of its canonical basis $\boldsymbol{E} = (\boldsymbol{e}_j)_{j\in[\![1,N]\!]}$ corresponds to a different z-index z_j throughout $\mathfrak{P}(x,y)$.

5.3. Data matrix decomposition

First, the sectional observations of $\boldsymbol{X}(x,y)$ are centered and their N-by-N covariance matrix $\boldsymbol{C_X}(x,y)$ is computed using Eq. 8. $\boldsymbol{C_X}(x,y)$ is then diagonalized using Eq. 9 so as to obtain its eigenvalues $(\lambda_{ij})_{i=j\in[\![1,N]\!]}$ in increasing order and its associated eigenvectors $\boldsymbol{G}(x,y) = (\boldsymbol{g}_j)_{j\in[\![1,N]\!]}$ of lengths N. The eigenvectors $(\boldsymbol{g}_j)_{j\in[\![1,N]\!]}$ of $\boldsymbol{C_X}(x,y)$ form a novel orthonormal basis for the cross-sectional responses of $\boldsymbol{X}(x,y)$. This eigenbasis $\boldsymbol{G}(x,y) = (\boldsymbol{g}_j)_{j\in[\![1,N]\!]}$ significantly describes them in a more discriminating way than their canonical basis $\boldsymbol{E} = (\boldsymbol{e}_j)_{j\in[\![1,N]\!]}$. Indeed, the first eigenvector $\boldsymbol{g}_1$ accounts for as much of the inertia $I(x,y)$ exhibited by the sectional observations of $\boldsymbol{X}(x,y)$ as possible. It indicates the direction of the Hilbert space $\mathbb{R}^N$ in which cross-sectional responses of $\boldsymbol{X}(x,y)$ exhibits the maximum variance λ_1. The succeeding eigenvectors, in the increasing order and subject to their mutual orthogonal condition, maximize their variances among the remaining inertia. Thus, most of the (interesting) information of the inertia $I(x,y)$ is considered by the largest eigenvectors (*i.e.* associated with the largest eigenvalues) (Jolliffe, 2002).

5.4. Dimension reduction

As the interesting inertia of the sectional observations of $\boldsymbol{X}(x,y)$ is considered by only the first few eigenvectors, the eigenbasis can be reduced to them, undergoing a dimensionality reduction (Jolliffe, 2002). The subset constituted by the first K eigenvectors $\boldsymbol{G}_K(x,y) = (\boldsymbol{g}_k)_{k\in[\![1,K]\!]}$ is thus regarded as a novel reduced eigenbasis, where K is a parameter whose value adjustment will be discussed in the subsection 5.6. Let $I_{\boldsymbol{G}_K}(x,y)$ denotes the inertia of the sectional observations of $\boldsymbol{X}(x,y)$ onto this reduced eigenbasis $\boldsymbol{G}_K(x,y)$: $I_{\boldsymbol{G}_K}(x,y) = \sum_{k=1}^{K} \lambda_{kk}(x,y)$.

5.5. Focus measurement

Because of the low-pass filter action of the PSF, a defocused part of the object surface $\mathcal{S}$ appears less contrasted on the acquired image than when it is focused: its corresponding pixels exhibit closer intensities, whose variance is thus lower (Sugimoto and Ichioka, 1985). Along an orthographic projection ray $\mathfrak{P}$, the variances of the sectional observations obviously match with the variation of their sharpness.

First let us introduce the measurement $\lambda\|\langle\boldsymbol{g},\boldsymbol{e}\rangle\boldsymbol{e}\|$ that is the norm of the orthogonal projection of the eigenvector $\boldsymbol{g}$ scaled by its respective eigenvalue λ onto the canonical vector $\boldsymbol{e}$. It gauges the amount retained by $\boldsymbol{e}$ among the variance λ accounted for by $\boldsymbol{g}$. By regarding the novel reduced basis $\boldsymbol{G}_K(x,y) = (\boldsymbol{g}_k)_{k\in[\![1,K]\!]}$ of $\mathfrak{P}(x,y)$, this amount (therefore among the inertia $I_{\boldsymbol{G}_K}(x,y)$) for a particular canonical vector $\boldsymbol{e}_p$ (with $p \in [\![1,N]\!]$) then appears as a possible estimation of the degree of focus of its corresponding voxel (x,y,z_p) of index z_p. Throughout $\mathfrak{P}(x,y)$, we propose as focus measurement the sum of the orthogonal projection norms of the eigenvectors making the novel reduced basis $\boldsymbol{G}_K(x,y) = (\boldsymbol{g}_k)_{k\in[\![1,K]\!]}$ and scaled by their respective eigenvalues $(\lambda_{kk})_{k\in[\![1,K]\!]}$ onto each canonical vectors $\boldsymbol{e}_p$, yielding the 3-D focus degree measure $\mathcal{F}$ as follows:

$$\mathcal{F}(x,y,z_p) = \sum_{k=1}^{K} \lambda_{kk}(x,y)\|\langle\boldsymbol{g}_k(x,y),\boldsymbol{e}_p\rangle\boldsymbol{e}_p\|. \tag{15}$$

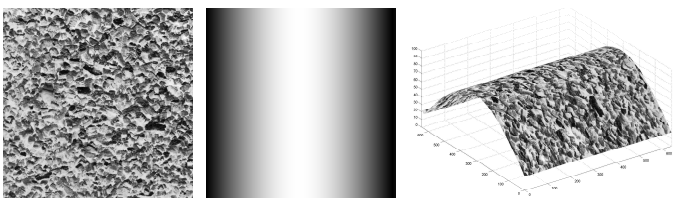


Figure 4: Generation of a simulated data sequence containing 32 images: (from left to right) D5 classical Brodtaz texture (Brodatz, 1966), fold-shaped artificial depth map (ground truth) and the corresponding 3-D reconstructed surface.

r (in pixels)	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2-D VAR	0.66	0.56	0.47	0.40	0.44	0.40	0.40	0.43	0.46	0.49	0.52	0.56	0.59	0.62
2-D TEN	3.87	2.41	1.48	0.93	0.59	0.47	0.44	0.45	0.48	0.51	0.55	0.59	0.63	0.68
2-D OPT	2.19	1.55	1.28	1.09	0.95	0.79	0.66	0.51	0.41	0.39	0.40	0.42	0.45	0.48
3-D DCT-PCA	6.47	7.63	8.31	8	7.14	6.06	5.11	4.62	4.32	4.26	4.35	4.33	4.21	4.26
3-D EIG K=1	1.37	1.39	1.26	1.03	0.77	0.63	0.54	0.47	0.47	0.49	0.52	0.54	0.57	0.59
3-D EIG K=10	0.86	0.79	0.62	0.59	0.49	0.41	0.40	0.43	0.46	0.49	0.53	0.56	0.60	0.64
2.5-D DEC	1.14													

Table 2: Performances (RMSE) of the SFF/EDF methods for the simulated data of Fig. 4, as a function of the neighborhood size r for the focus-based ones (unlike the model-based 2.5-D DEC one). In absence of noise, the performances of our 3-D EIG method are close to the others.

5.6. Noise robustness vs. focus resolution

In this subsection, the adjustment of the parameter K reducing the considered eigenbasis $\boldsymbol{G}_K$ is discussed. As previously mentioned, noises/disturbances result in incorrect information of sharpness (*i.e.* they generally constitute high-spatial-frequency components of an image, like focus cues) that misleads the reconstruction process. Nevertheless, this problem can be minimized in our proposed 3-D EIG analysis by eliminating the lower range of eigenvectors. Throughout the 3-D image $\boldsymbol{\mathcal{I}}$, a (well-contrasted) correct data has a larger influence in the eigenvalues decomposition than a noisy/disturbance one. Even if it sectionally appears more contrasted, this latter generally affects a narrower range of depth than the one of interest. As a result, its respective information is restricted to lower order eigenvectors (Fukunaga, 1990) and a small value of the parameter K allows to reduce it in the focus measure. The lower the K parameter value is, the larger the noise robustness of our 3-D EIG method is.

In the same way, such a small value of K might, on the contrary, disregard information of a correct but poorly contrasted data. In opposition, a larger value of K increases the focus measurement sensitivity: the 3-D EIG measurement gains in focus resolution but loses robustness to noises/disturbances; the focus resolution being the capacity of a focus measurement to distinguish the poorest contrasted but focused data. So, there is a trade-off associated with choosing the appropriate value of the K parameter. However, a K value set to 1 will be preferred in most usual cases, moreover necessarily tarnished by noise (Fernandes et al., 2011b). The eigenvalues decomposition intrinsically results in eigenvalues $(\lambda_{ij})_{i=j\in[\![1,N]\!]}$ that rapidly decreases and whose the first one λ_1 accounts for the major part of the total inertia I (Jolliffe, 2002).

In addition, the parameter K can be locally adjusted in an adaptive way throughout the input sequence $\boldsymbol{\mathcal{I}}$. For each orthographic ray $\mathfrak{P}(x,y)$, $K(x,y)$ can be chosen so that, *e.g.* the normalized inertia $\widetilde{I}_{\boldsymbol{G}_K}$ exhibited onto the reduced eigenbasis $\boldsymbol{G}_K(x,y) = (\boldsymbol{g}_j)_{j\in[\![1,K]\!]}$:

$$\widehat{I}_{\boldsymbol{G}_K}(x,y) = \frac{I_{\boldsymbol{G}_K}}{I(x,y)} = \sum_{k=1}^{K} \widehat{\lambda}_{kk}(x,y) = \sum_{k=1}^{K} \frac{\lambda_{kk}(x,y)}{\mathrm{tr}(\boldsymbol{\Lambda}(x,y))}, \qquad (16)$$

is under a fixed threshold $s \in [0,1]$, such as:

$$K(x,y) = \underset{K}{\mathrm{argmax}} \left\{ \widehat{I}_{\boldsymbol{G}_K}(x,y) \;/\; \widehat{I}_{\boldsymbol{G}_K}(x,y) \leqslant s \right\}. \qquad (17)$$

Thus, the threshold s may be set so as to consider, *e.g.* , the first 90 per cent of the inertia $I(x,y)$, excluding the last 10 per cent highly related to noisy content.

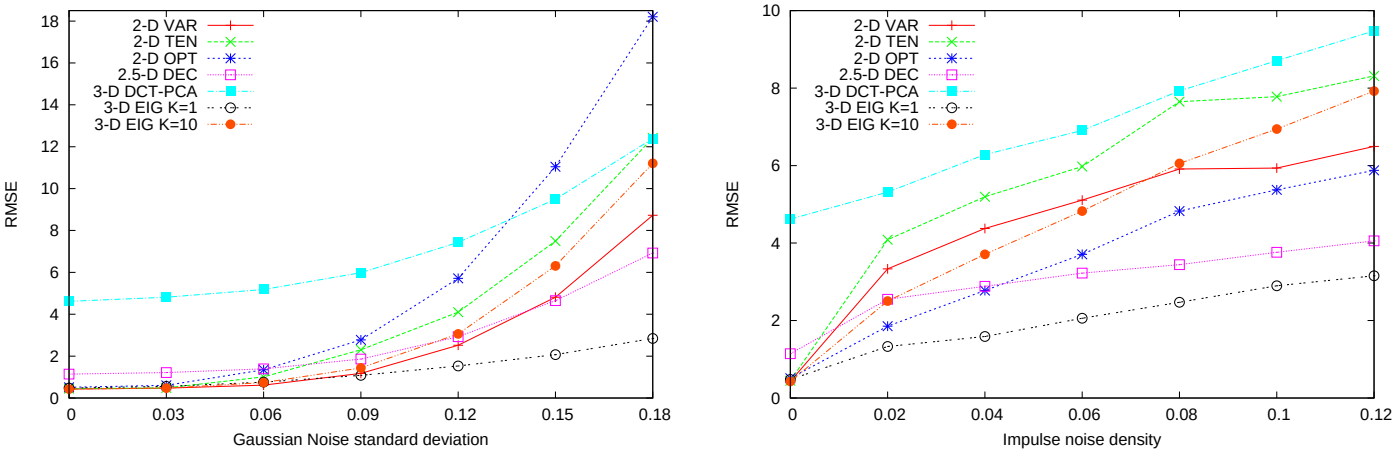
5.7. Computational aspects

Because of its 3-D strategy, the proposed 3-D EIG method incurs a higher computational cost than the 2-D reference ones. Both covariance matrix calculation and diagonalization contribute most of its computational complexity and require $\mathrm{O}(\sharp\mathfrak{P}^2\,\sharp\mathfrak{B}_r)$ and $\mathrm{O}(\sharp\mathfrak{P}^3)$ computations, respectively. Nevertheless, the only involved operations are matrix-vector multiplications, easily implementable. Furthermore, a number of methods have been proposed to reduce the computational complexity of the second diagonalization step. Notably, Sharma and Paliwal (2007) introduced a fixed-point algorithm, called Fast PCA, whose estimated computational complexity is of order $\mathrm{O}(\sharp\mathfrak{P}^2\,K)$. Finally, the process being carried out at each voxel stack $\mathfrak{P}$ independently, the computational cost can be substantially minimized by employing a parallel implementation of the 3-D EIG method, moreover with a GPU (Graphics Processing Units) parallel optimized version (Andrecut, 2009).

6. Validation on synthetic data

In order to dispose of a ground truth for carrying out quantitative assessments of the results produced by the state-of-the-art retained methods (summarized at Tab. 1) and our 3-D EIG approach, a first experiment using synthetic data is conducted. So, the theoretical robustness to several artificial additive noises will be particularly studied.

By first mapping an arbitrary texture onto an artificial depth map (that constitutes the ground truth), a synthetic 3-D surface is constructed (Fig. 4). Notice that this surface topography is relatively smooth, consistent with the ones encountered in the dealt-with medical issue (see the next Section 7). Afterwards, a sequence of 2-D images is generated and collected by using a discrete version of Eq. 1 for a finite range of successive locations of the object focal plane by constant step sizes Δz. To that end, the 2-D PSF is approximated by a 2-D Gaussian function normalized to account for an uniform illumination (*e.g.* a Köhler illumination), whose standard deviation is proportional to the distance of defocus δz (Aguet et al., 2008). Additive Gaussian or impulse noise are finally introduced to the

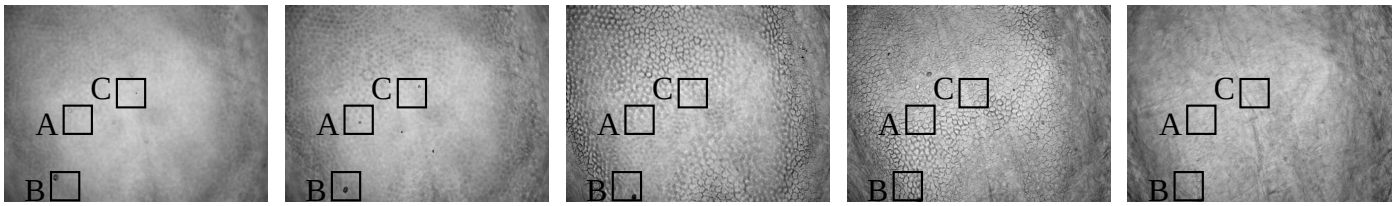


(a) With Gaussian noise (b) With impulse noise

Figure 5: Performances (RMSE) of SFF/EDF methods (if necessary, $r = 8$ pixels) for the simulated data of Fig. 4 under various noisy conditions. Our 3-D EIG method clearly outperforms the other ones in presence of noise.

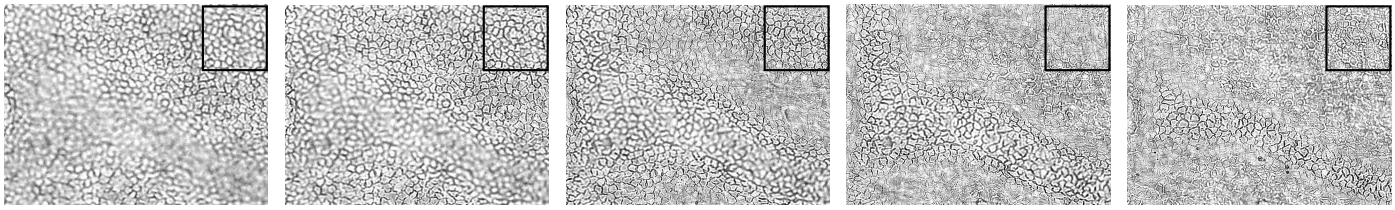
simulated image sequence to test the robustness of the aforementioned methods. Performances are measured in terms of the root-mean-square-error (RMSE) metric with respect to the ground truth (Gonzalez and Woods, 2008).

Tab. 2 reports the performances of aforementioned approaches on the synthetic image sequence of Fig. 4 without noise, and as a function of the neighborhood size r for the focus measurements (*i.e.* 2-D VAR, 2-D TEN, 2-D OPT, 3-D DCT-PCA and 3-D EIG methods). In view of the fact that the synthetic surface exhibits neither sharp depth slopes nor discontinuities, these ones perform better with a rather large size r. In absence of noise, the performance of the 3-D EIG method is similar to the other ones, despite the performed 3-D strategy that tends to slightly enlarge the focus resolution. As expected, remark that a larger value of K improves its performances for small sizes r by narrowing its focus resolution needing to distinguish some critical texture regions, *e.g.* wide homogeneous regions made up of few high-spatial-frequency cues. As for the



(a) Section 1 (b) Section 10 (c) Section 20 (d) Section 30 (e) Section 40

Figure 6: Some individual 2-D image sections among the 40 composing the image sequence 1. This sequence was imaged in steps of 4.5 μm using an Olympus BX 41 transmitted white-light microscope equipped with a ×10 / NA 0.25 objective in air immersion. Each image section is 1040×772 pixels, representing 718×533 μm. Note that both bottom left corner and right edge regions never appear in-focus throughout the sequence. The marked regions A, B and C will be used as sites for comparison of restored textures. The cellular fragments dumped in the dilating biochemical solution are clearly visible on (a) and (b) as small dark spots, *e.g.* throughout the region B.



(a) Section 17 (b) Section 21 (c) Section 25 (d) Section 29 (e) Section 33

Figure 7: Some individual 2-D image sections among the 35 composing the image sequence 2. This sequence was imaged in steps of 4.5 μm using a Leica Laborlux S transmitted white-light microscope equipped with a ×10 / NA 0.25 objective in air immersion. Each image section is 759×574 pixels, representing 759×574 μm. The marked region will be used as a site for comparison of restored textures. The contrast reversals are clearly visible (see highlights of the marked region in Fig. 9).

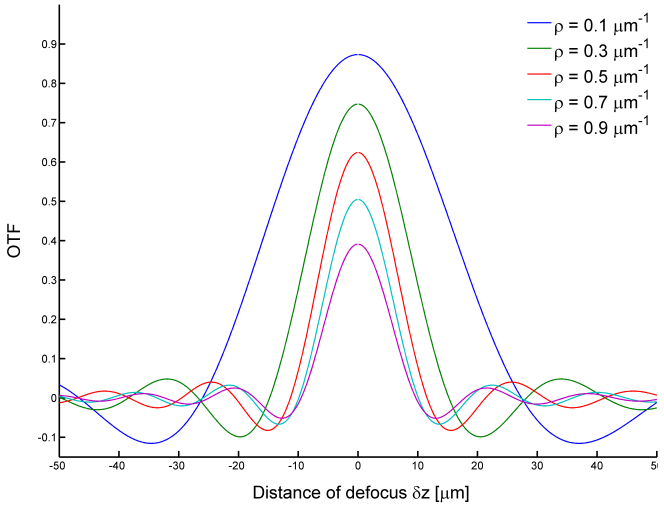


Figure 8: Castleman's OTF approximation (Castleman, 1996) of a simulated ×10 / NA 0.25, WD 10.5 mm objective as a function of the distance of defocus δz and for some radial components $\rho = \sqrt{\nu_x^2 + \nu_y^2}$ of both spatial-frequency coordinates ν_x and ν_y. The OTF correctly acts as a low-pass filter whose cut-off frequency decreases when the distance of defocus δz increases. Particularly note that the OTF, out of a certain degree of defocus, exhibits negative resonances for some spatial-frequency ranges that are thus subject to contrast reversals in the acquired image. As δz increases, the OTF becomes increasingly oscillatory, with the contrast reversals affecting ever smaller spatial-frequencies. Finally, remark the slight OTF asymmetry from the focus position.

other 3-D focus measurement (3-D DCT-PCA), it shows the weakest performances: the previous discrete cosine transformation does not allow the statistical analysis to be sufficiently sensitive and discriminating.

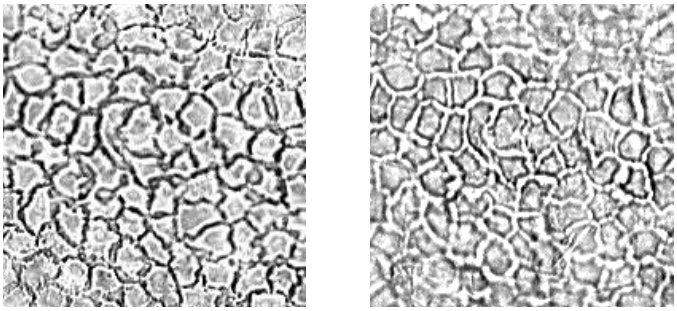
Secondly, SFF/EDF methods are put to the test (still on the simulated image sequence of Fig. 4) with various additive noises: their performances are shown in Fig. 5. According to previous noise-free experiment, the size r of neighborhood-based focus measurements is here approximately adjusted to the trade-off value of 8 pixels (see Tab. 2). In presence of noise, the proposed 3-D EIG method with K set to 1 clearly outperforms the other ones. This robustness to noise is due to the adopted 3-D statistical strategy: the simultaneous use of all cross-sectional cues throughout the image sequence makes the discrimination of in-focus positions "drowned" in noise possible. As expected, the lower the K parameter value is, the more robust to noise the 3-D EIG method is.

7. Application to 3-D reconstruction of human *ex-vivo* corneal endotheliums

We now illustrate the potential of our 3-D EIG method on real acquisitions from the dealt-with medical issue: the 3-D reconstruction of endothelial surfaces of human corneal grafts, in order to improve their quality control.

7.1. Disturbed image acquisitions

Endotheliums of human corneal grafts are inspected by observing the grafts from the backside using conventional optical transmitted light microscopy. Two sequences tagged 1 and 2 are

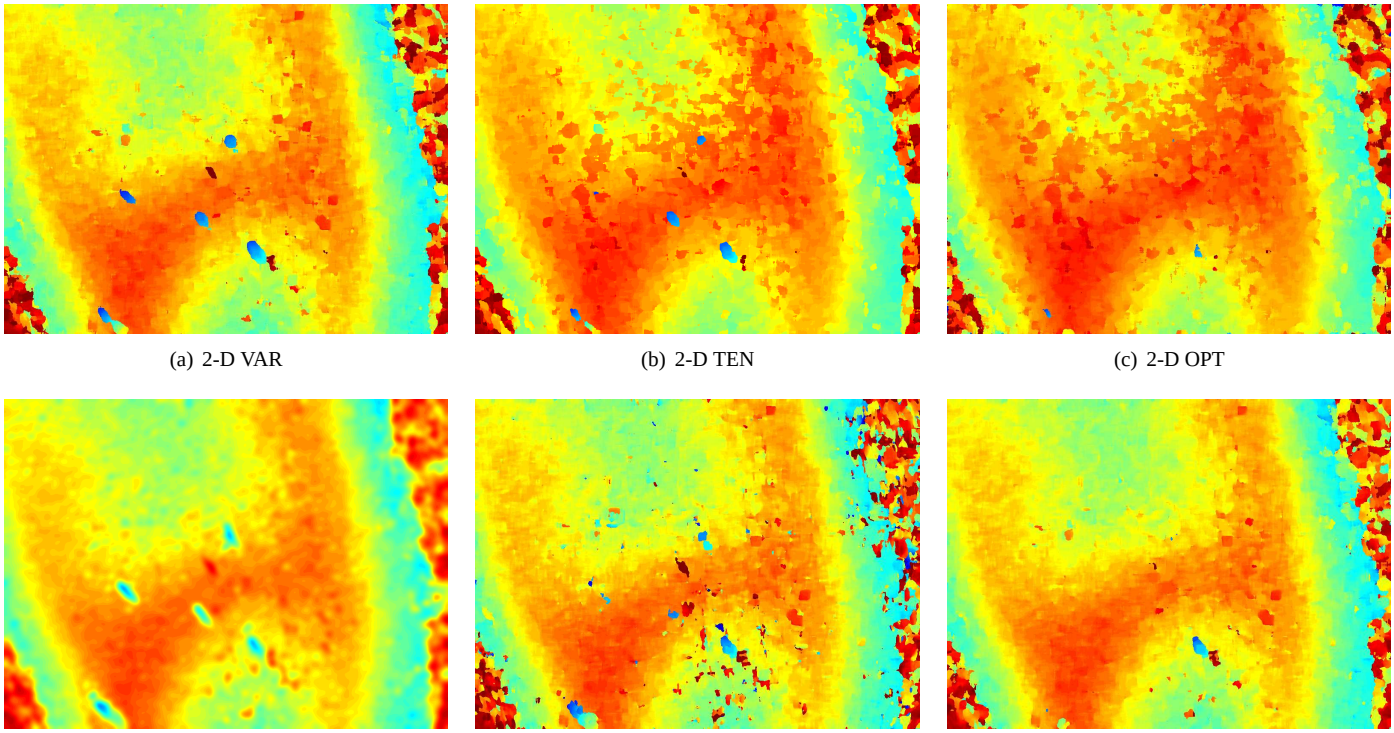


(a) Near-focus section 25 (b) Out-of-focus section 33

Figure 9: Illustration of contrast reversals affecting both endothelial cell rims and borders by zooming in on two regions of the image sequence 2 of Fig. 7.

considered herein; both are described and illustrated in Fig. 6 and Fig. 7 respectively. The graft concavity is beforehand filled with a specific biochemical solution that dilates the intercellular spaces between endothelial cell bodies (referred to as cell borders) (Sperling, 1986). Thereby, the cell borders are more opaque than the whole bodies and normally appear darker (remark that they are often poorly visible). SFF/EDF concepts appear here as adapted ways so as to reconstruct the endothelial scene surfaces. Indeed, through such conventional optical microscope, the inherent use of large magnifications $\gamma \sim$ NA significantly limits the offered depth-of-field and the performed projection tends towards an orthographic behaviour (and therefore an all-over constant magnification) since the working distance WD $= |z_o|$ is much larger than the profile thickness of the observed surface. Nevertheless, such observations necessarily result in noisy and disturbed acquisitions. For example, sequence 1 is disturbed by some cellular fragments dumped in the dilating solution (see Fig. 6). Both are even more damaged by intense contrast reversals (see Fig. 6, and more deeply, Fig. 7 and its highlights in Fig. 9). This disturbance data throughout the image sequence misleads the reconstruction process, preventing from practical results.

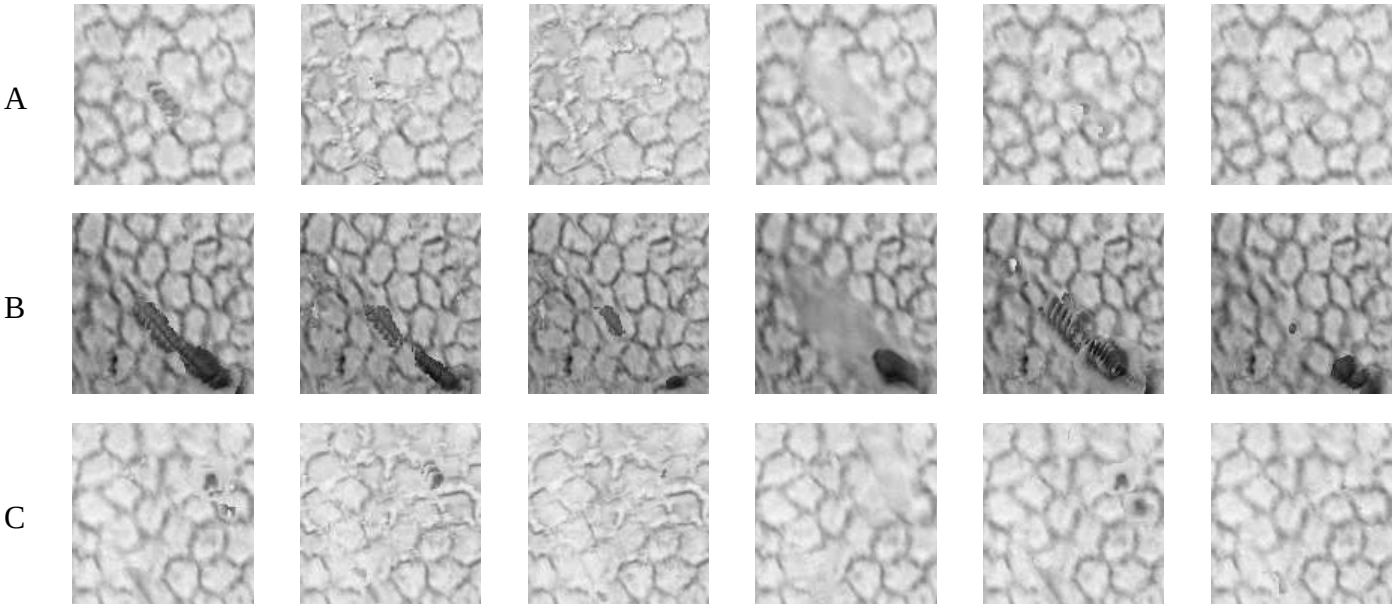
The PSF can effectively cause contrast reversals for some size ranges of textural contents, *i.e.* from bright to dark or vice-versa between the observed scene surface and the acquired image (Mahajan, 2001). In order to briefly investigate this phenomenon, it would be appropriate to introduce the optical transfer function (OTF) that is the Fourier transform of the PSF $\boldsymbol{h}$ and therefore transforms the blurring operation of Eq. 1 and Eq. 2 from convolution to multiplication. An OTF approximation according to Castleman's formulation (Castleman, 1996) is depicted in Fig. 8. Effectively, the OTF becomes negative for some distinct ranges of distances of defocus δz, then causing contrast reversals for some specific ranges of spatial-frequency components of the surface texture. This is the case for the endothelial cell borders that are normally darker than the cell bodies and look brighter on several distinct 2-D image sections of the sequence (see Fig. 9). Remark that contrast reversals are here rather emphasized and intense, notably because of both sample over-illumination and transmitted light imaging. In many cases, the cell body rims are extremely bright (and satu-



(a) 2-D VAR (b) 2-D TEN (c) 2-D OPT

(d) 2.5-D DEC (e) 3-D DCT-PCA (f) 3-D EIG $K = 1$

Figure 10: Reconstructed depth maps for the image sequence 1. The color z-scale is: **0** • • 29.25 • • 58.5 • • 87.75 • • 117 • • 146.25 • • 175.5 μm. The 3-D EIG depth map (f) distinctly contains fewer both blue spots and over-red regions caused by cellular fragments and contrast reversals, respectively.



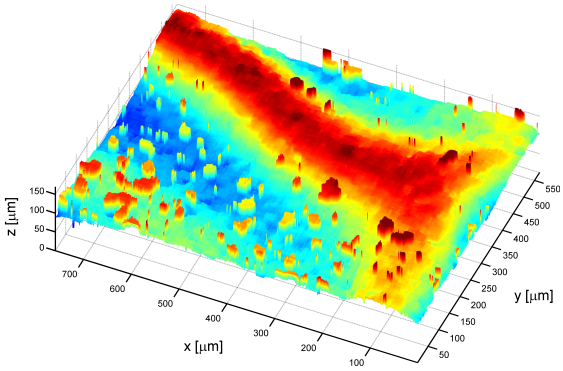
(a) 2-D VAR (b) 2-D TEN (c) 2-D OPT (d) 2.5-D DEC (e) 3-D DCT-PCA (f) 3-D EIG $K = 1$

Figure 11: Details of the restored textures for the image sequence 1 in the regions A, B, C of Fig. 6. Artifacts attributed to both cellular fragments and contrast reversals are noticeably fewer on the details of the texture in (f) restored by the 3-D EIG method.

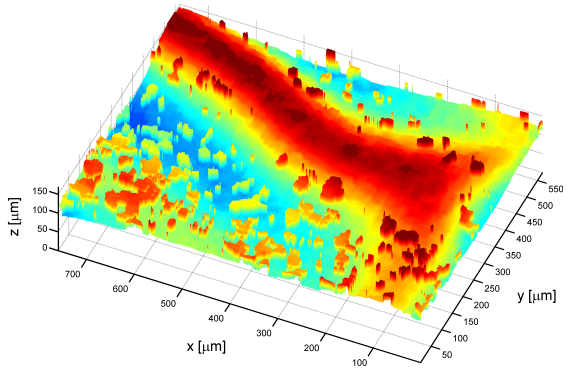
rated at the focus on the acquired image). At contrast reversals occurrence, they then lead to saturated bright cell borders on the acquired image and, as for them also affected, become darker but narrower.

7.2. *Results and discussion*

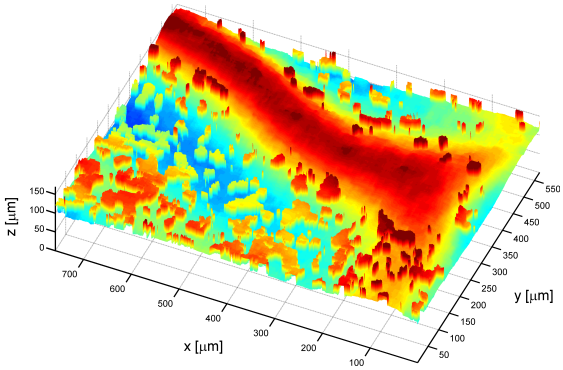
We now discuss the results of state-of-the-art retained methods (summarized at Tab. 1) and our 3-D EIG approach on the couple of image sequences 1 and 2. Given their noisy and dis-



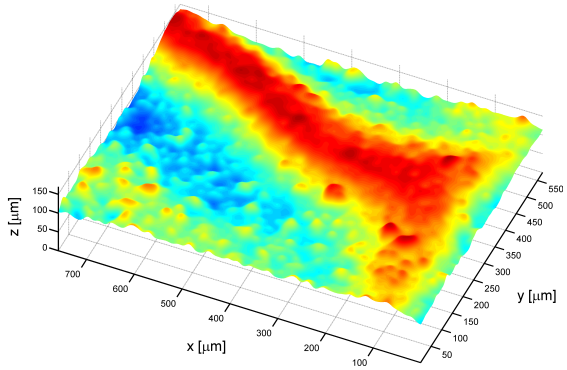
(a) 2-D VAR



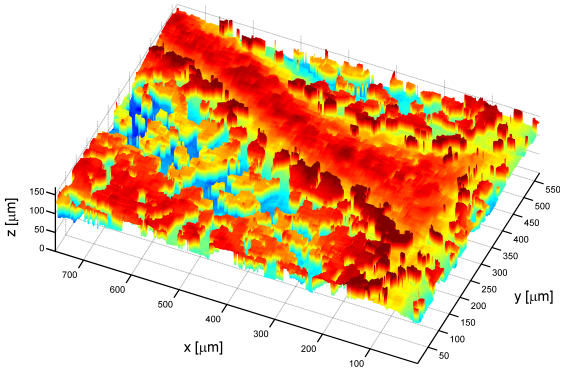
(b) 2-D TEN



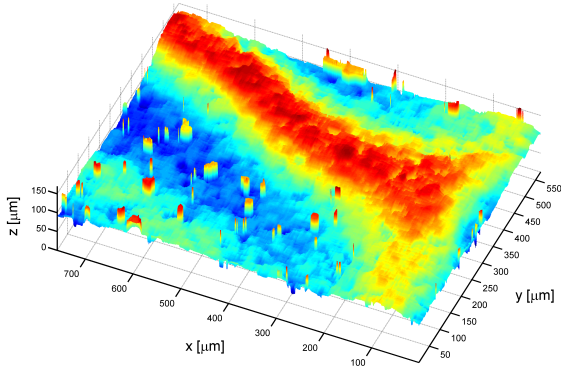
(c) 2-D OPT



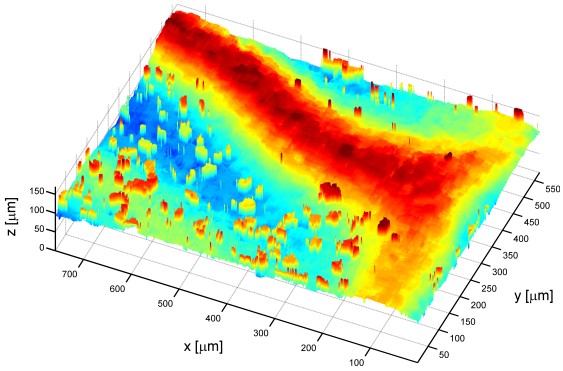
(d) 2.5-D DEC



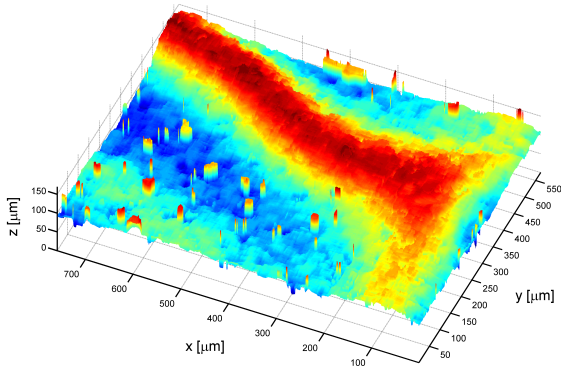
(e) 3-D DCT-PCA



(f) 3-D EIG $K = 1$

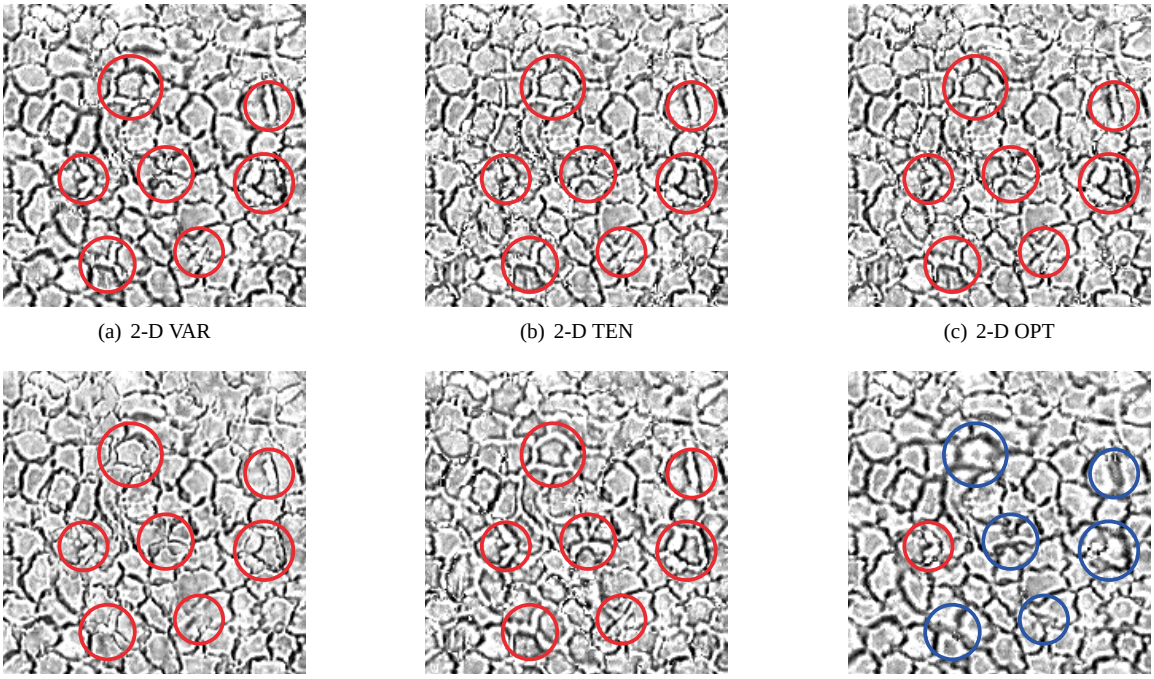


(g) 3-D EIG $K = 10$



(h) K-adaptive 3-D EIG with $s = 0.55$

Figure 12: Reconstructed depth maps for the image sequence 2. The color z-scale is: 75 • • • 88 • • • 101 • • • 114 • • • 127 • • • 140 • • • 153 μm. The 3-D EIG depth maps (f-h) distinctly contain fewer over-red spike regions caused by intense contrast reversals.



(a) 2-D VAR (b) 2-D TEN (c) 2-D OPT

(d) 2.5-D DEC (e) 3-D DCT-PCA (f) K-adaptive 3-D EIG with $s = 0.55$

Figure 13: Details of the restored textures for the image sequence 2 in the marked region of Fig. 7. Red circles highlight some mis-recovered regions because of the presence of contrast reversals, whereas the well-reconstructed ones indicated by blue circles. The reader can refer to the original images in Fig. 9 for comparison.

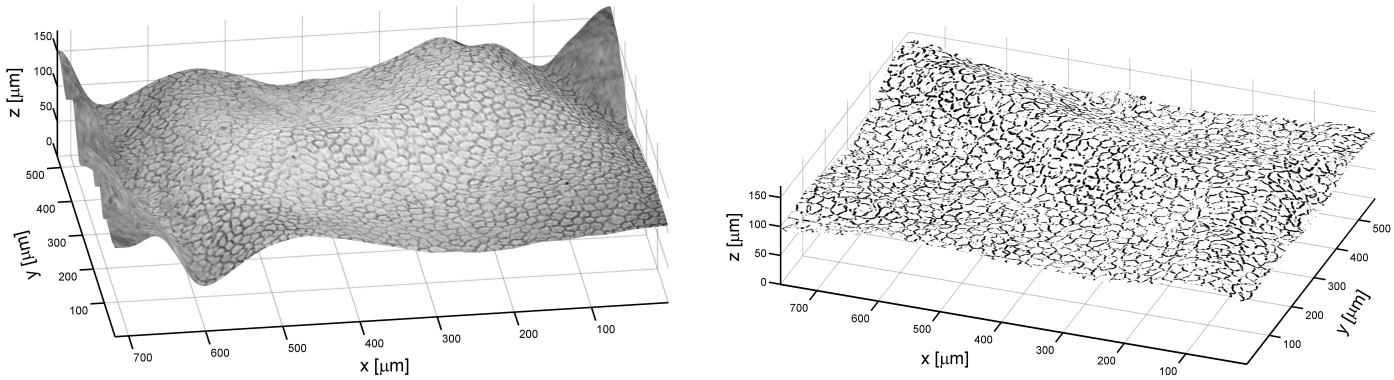
turbed appearances, only a robust SFF/EDF method will succeed in satisfactory reconstructing the endothelial surfaces. The neighborhood-based focus measurements are performed with the large size r of 10 pixels, because of both wide textural content and noisy aspect of the endothelial surfaces. Moreover, this is here not prejudicial in view of the relative depth regularity of the surface (see the results on the synthetic data shown in Tab. 2). The reconstructed depth maps are shown in Fig. 10 and Fig. 12; the restored textures are highlighted in Fig. 11 and Fig. 13.

Depth maps inferred by the proposed 3-D EIG method with K set to 1 clearly exhibit fewer artefacts, anatomically impossible as the endothelium surface is necessarily continuous. In Fig. 10, it distinctly contains fewer underestimated (blue) and overestimated (over-red) regions caused by cellular fragments and contrast reversals, respectively. In Fig. 12, it clearly exhibits fewer overestimated (over-red) spike regions and discontinuities due to the presence of intense contrast reversals throughout the sequence. As for the model-based 2.5-D DEC method, resulting depth maps and the more numerous artefacts damaging them appear smoothed owing to the performed coarse-to-fine optimization (see Subsection 3.2).

In some regions, the 3-D EIG $K = 1$ method slightly lacks focus resolution and therefore accuracy, made visible when no disturbance arises, *e.g.* at the top of the fold of the sequence 2 (Fig. 12(f)). As expected, a larger value of K, *e.g.* set to 10, increases the focus resolution but instead decreases the robustness (Fig. 12(g)). In this case, an adaptive parameter K function of the presence or not of contrast reversals should improve the surface reconstruction. If contrast reversals are present, their information is normally restricted to the second eigenvector $\boldsymbol{g_2}$, which then accounts for a much larger normalized eigenvalue $\widehat{\lambda}_2$. Thereby, the normalized inertia $\widehat{I}$ is mainly shared between both $\widehat{\lambda}_1$ (contrast normal sense) and $\widehat{\lambda}_2$ (contrast reversal). Otherwise, the first eigenvector $\boldsymbol{g_1}$ accounts for the major part *via* $\widehat{\lambda}_1$. Consequently, the adaptive $K(x, y)$ can be simply adjusted to 1 if $\widehat{\lambda}_1(x, y)$ is under a certain threshold s or a much larger value, *e.g.* 10, otherwise. In this way, the 3-D EIG method offers a robust behaviour in presence of contrast reversals, whereas it provides a more sensitive focus resolution in all other cases (see Fig. 12(h)).

Restored textures are respectively highlighted on some crucial regions in Figs. 11 and 13 for a better visibility. Their inspection corroborates the above appreciation, as each of them being intimately related to the associated depth map. So, the 3-D EIG method achieves strong improvements: its restored textures are not too much damaged by disturbances, like cellular fragments or contrast reversals.

To cope with noisy/disturbance issues, the 3-D EIG method exploits the fact that they generally affect a narrower range of depth than the one of interest (the true surface), in spite of sometimes exhibiting the largest contrasts. This notice is only possible through a 3-D strategy within the image sequence. Here, the cellular fragments move continually and the OTF negative resonance causing the contrast reversals theoretically reveals a narrower bandwidth (Fig. 8). Consequently, noisy/disturbance data has a lesser influence in the eigenvalues decomposition than the one of interest and their respective information is pushed into



(a) 3-D EIG $K = 1$ (b) 3-D EIG $K = 1$

Figure 14: 3-D reconstructions of the endothelial surfaces for both image sequences 1 and 2 using the 3-D EIG method. The posterior folds are clearly visible and the endothelial cell mosaics appear sharply as a whole, except for some regions of (a) never in-focus throughout the image sequence 1 (see Fig. 6).

lower order eigenvectors.

Finally, 3-D EIG depth maps are interpolated and smoothed using cubic B-splines, respective textures are deduced and 3-D reconstructions of the surfaces are recovered by mapping texture images onto depth maps. Three-dimensional endothelial surface reconstructions for both image sequences 1 and 2 are shown in Fig. 14, thereby making the examination of the endothelial cell mosaics over wider regions possible.

8. Conclusions

In this paper, we have introduced a new method designated as 3-D EIG for 3-D image reconstructions of scene surfaces from optical sectioning with a limited depth-of-field imaging system. Through the common SFF/EDF concepts, the suggested 3-D EIG method succeeds in recovering both depth map and texture of the observed surface, without needing any knowledge of the characteristics of the imaging system. By locally performing multivariate statistical analyses within cross-sectional spatial windows, it originally exploits all cross-sectional (axial) cues simultaneously to infer the observed surface throughout the image sequence, whereas state-of-the-art methods generally follow a sectional way. This adopted 3-D statistical strategy offers a strong robustness that is an important constraint of focus analysis and SFF/EDF concepts, more particularly when dealing with noisy and disturbed acquisitions as in the addressed medical issue concerning the 3-D reconstructions of human organ-cultured corneal endotheliums.

In the proposed approach, the K parameter enables to gauge the noise robustness versus the focus resolution. Indeed, a small value of K confers a large robustness but causes a slight enlargement of the focus resolution. Notice that this K parameter value can be automatically and adaptively adjusted within the image sequence, *e.g.* by simply imposing a threshold on the normalized eigenvalues. As for the choice of the neighborhood size parameter r, there is a trade-off too. Large values raise the robustness but reduce the spatial resolution and increase the computational time. On the contrary, smaller sizes increase the spatial resolution and decrease the computational time but are more affected by noise and homogeneous textural regions.

The validity of our 3-D EIG method is clearly demonstrated on both synthetic data and real acquisitions of the addressed medical issue. The simulation experiment shows that without noise the 3-D EIG approach exhibits performances close to state-of-the-art methods, while it outperforms them in presence of noise. Real acquisitions of organ-cultured corneal endotheliums present difficult noisy/disturbance problems, for which the improvements on account of the manifested robustness are clearly visible. Our 3-D EIG approach succeeds in reconstructing the endothelial surfaces of human corneal grafts, making the examination of the cell mosaics over wider sharp regions possible. We conceive automatically estimating endothelial cell density in 2-D on the restored texture image (Gavet and Pinoli, 2008) and then correct it through topographical information of the inferred depth map, yielding a 3-D estimation.

Despite the relative regularity of the studied surfaces, it is important to stress the fact that the 3-D-EIG approach is not limited to any specific kind or model of surfaces (Fernandes et al., 2011b). Furthermore, while it is illustrated in the context of conventional optical microscopy, it is applicable to the wider range of imaging systems offering a limited depth-of-field, provided the acquired image sequence is previously registered.

Acknowledgments

The authors wish to thank Pr. Gilles Thuret and Pr. Philippe Gain from University Hospital Centre, Saint-Étienne, France for supporting this work and for providing pictures of human organ-cultured corneal endotheliums.

References

Agard, D. A., 1984. Optical sectioning microscopy - Cellular architecture in three dimensions. Annual Review of Biophysics and Bioengineering 13, 191–219.

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

Ahmad, M., Choi, T.-S., 2005. A heuristic approach for finding best focused shape. IEEE Transactions on Circuits and Systems for Video Technology 15 (4), 566–574.

Andrecut, M., 2009. Parallel GPU implementation of iterative PCA algorithm. Journal of Computational Biology 16 (11), 1593–1599.

Asif, M., Choi, T.-S., 2001. Shape from focus using multilayer feedforward neural networks. IEEE Transactions on Image Processing 10 (11), 1670–1675.

Born, M., Wolf, E., 1991. Principles of optics - Electromagnetic theory of propagation interference and diffraction of light, 6th Edition. Pergamon Press, New York, USA.

Brenner, J. F., Dew, B. S., Horton, J. B., King, T., Neurath, P. W., Selles, W. D., 1976. Automated microscope for cytologic research: Preliminary evaluation. Journal of Histochemistry & Cytochemistry 24 (1), 100–111.

Brodatz, P., 1966. Textures: a photographic album for artists and designers. Dover Publications, New York, USA.

Burt, P., Adelson, E., 1983. The laplacian pyramid as a compact image code. IEEE Transactions on Communications 31, 532–540.

Castleman, K. R., 1996. Digital image processing. Prentice Hall, Upper Saddle River, NJ, USA.

Darrell, T., Wohn, K., 1988. Pyramid based depth from focus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Ann Arbor, MI , USA, pp. 504–509.

Fernandes, M., Gavet, Y., Pinoli, J.-C., 2011a. Improving focus measurements using logarithmic image processing. Journal of Microscopy 242 (3), 228–241.

Fernandes, M., Gavet, Y., Pinoli, J.-C., 2011b. Robust shape-from-focus by 3-D multivariate statistical analyses. In: Proceedings of the IEEE International Conference on Image Processing. Brussels, Belgium, pp. 2113–2116.

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

Frieden, B. R., 1967. Optical transfer of the three-dimensional object. Journal of the Optical Society of America 57 (1), 56–65.

Fukunaga, K., 1990. Introduction to statistical pattern recognition, 2nd Edition. Academic Press, San Diego, CA, USA.

Gain, P., Thuret, G., Kodjikian, L., Gavet, Y., Turc, P. H., Theillere, C., Acquart, S., Le petit, J. C., Maugery, J., Campos, L., 2002. Automated tri-image analysis of stored corneal endothelium. British Journal of Ophthalmology 86 (7), 801–808.

Gavet, Y., Pinoli, J.-C., 2008. Visual perception based automatic recognition of cell mosaics in human corneal endothelium microscopy images. Image Analysis and Stereology 27, 53–61.

Gibson, S. F., Lanni, F., 1989. Diffraction by a circular aperture as a model for three-dimensional optical microscopy. Journal of the Optical Society of America A 6 (9), 1357–1367.

Gonzalez, R. C., Woods, R. E., 2008. Digital image processing, 3rd Edition. Prentice-Hall, Upper Saddle River, NJ, USA.

Horn, B., 2001. Robot vision, 13th Edition. The Massachusetts Institute of Technology Press, Cambridge, MA, USA.

Jolliffe, I., 2002. Principal component analysis, 2nd Edition. Springer-Verlag, New York, USA.

Krotkov, E., 1987. Focusing. International Journal of Computer Vision 1 (3), 223–237.

Mahajan, V. N., 1998. Optical imaging and aberrations part I: ray geometrical optics. SPIE Press, Bellingham, WA, USA.

Mahajan, V. N., 2001. Optical imaging and aberrations part II: wave diffraction optics. SPIE Press, Bellingham, WA, USA.

Mahmood, M. T., Choi, T. S., 2008. A feature analysis approach to estimate 3D shape from image focus. In: Proceedings of the IEEE International Conference on Image Processing. Vol. 1–5. San Diego, CA , USA, pp. 3216–3219.

Mahmood, M. T., Choi, W. J., Choi, T. S., 2008. PCA-based method for 3D shape recovery of microscopic objects from image focus using discrete cosine transform. Microscopy Research and Technique 71 (12), 897–907.

Mahmood, M. T., Shim, S. O., Choi, T. S., 2009. Shape from focus using principal component analysis in discrete wavelet transform. Optical Engineering 48 (5), 057203.

Malik, A. S., Choi, T. S., 2007. Consideration of illumination effects and optimization of window size for accurate calculation of depth map for 3D shape recovery. Pattern Recognition 40 (1), 154–170.

Malik, A. S., Choi, T. S., 2008. A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. Pattern Recognition 41 (7), 2200–2225.

Meneses, J., Suarez, M. A., Braga, J., Gharbi, T., 2008. Extended depth of field using shapelet-based image analysis. Applied Optics 47 (2), 169–178.

Nayar, S. K., Nakagawa, Y., 1994. Shape from focus. IEEE Transactions on Pattern Analysis and Machine Intelligence 16 (8), 824–831.

Pels, E., Schuchard, Y., 1983. Organ-culture preservation of human corneas. Documenta Ophthalmologica 56 (1-2), 147–153.

Pentland, A. P., 1987. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (4), 523–531.

Pieper, R. J., Korpel, A., 1983. Image processing for extended depth of field. Applied Optics 22 (10), 1449–1453.

Pradeep, K. S., Rajagopalan, A. N., 2007. Improving shape from focus using defocus cue. IEEE Transactions on Image Processing 16 (7), 1920–1925.

Sharma, A., Paliwal, K. K., 2007. Fast principal component analysis using fixed-point algorithm. Pattern Recognition Letters 28 (10), 1151–1155.

Sperling, S., 1986. Evaluation of the endothelium of human donor corneas by induced dilation of intercellular spaces and trypan blue. Graefes Archive for Clinical and Experimental Ophthalmology 224 (5), 428–434.

Strang, G., 1980. Linear algebra and its applications, 2nd Edition. Academic Press, San Diego, CA, USA.

Streibl, N., 1984. Depth transfer by an imaging-system. Optica Acta 31 (11), 1233–1241.

Subbarao, M., Choi, T., 1995. Accurate recovery of three-dimensional shape from image focus. IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (3), 266–274.

Subbarao, M., Choi, T., Nikzad, A., 1993. Focusing techniques. Optical Engineering 32 (11), 2824–2836.

Sugimoto, S. A., Ichioka, Y., 1985. Digital composition of images with increased depth of focus considering depth information. Applied Optics 24 (14), 2076–2080.

Thuret, G., Manissolle, C., Acquart, S., Garraud, O., Campos-Guyotat, L., Maugery, J., Gain, P., 2004. Urgent need for normalization of corneal graft quality controls in French eye banks. Transplantation 78 (9), 1299–1302.

Thuret, G., Manissolle, C., Acquart, S., Le Petit, J. C., Maugery, J., Campos-Guyotat, L., Doughty, M. J., Gain, P., 2003. Is manual counting of corneal endothelial cell density in eye banks still acceptable? The French experience. British Journal of Ophthalmology 87 (12), 1481–1486.

Toet, A., 1989. Image fusion by a ratio of low-pass pyramid. Pattern Recognition Letters 9 (4), 245–253.

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

Vollath, D., 1987. Automatic focusing by correlative methods. Journal of Microscopy 147, 279–288.

Watanabe, M., Nayar, S. K., 1997. Telecentric optics for focus analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (12), 1360–1365.

Wee, C. Y., Paramesran, R., 2007. Measure of image sharpness using eigenvalues. Information Sciences 177 (12), 2533–2552.

Willson, R. G., 1994. Modeling and calibration of automated zoom lenses. Tech. Rep. CMU-RI-TR-94-03, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA.

Willson, R. G., Shafer, S. A., 1991. Dynamic lens compensation for active color imaging and constant magnification focusing. Tech. Rep. CMU-RI-TR-91-26, The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA.

Yun, J., Choi, T., 1999. Accurate 3-D shape recovery using curved window focus measure. In: IEEE International Conference on Image Processing. Vol. 3. pp. 910–914 vol.3.

Zhang, Z., Blum, R. S., 1999. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. Proceedings of the IEEE 87 (8), 1315–1326.