

# Journal of Biomedical Optics

[SPIDigitalLibrary.org/jbo](http://SPIDigitalLibrary.org/jbo)

## **Anisotropy-based robust focus measure for non-mydriatric retinal imaging**

Andrés G. Marrugo  
María S. Millán  
Gabriel Cristóbal  
Salvador Gabarda  
Héctor C. Abril

# Anisotropy-based robust focus measure for non-mydratiac retinal imaging

Andrés G. Marrugo,<sup>a</sup> María S. Millán,<sup>a</sup> Gabriel Cristóbal,<sup>b</sup> Salvador Gabarda,<sup>b</sup> and Héctor C. Abril<sup>a</sup>

<sup>a</sup>Universitat Politècnica de Catalunya, Group of Applied Optics and Image Processing, Department of Optics and Optometry, Violinista Vellsolà 37, 08222 Terrassa, Spain

<sup>b</sup>Instituto de Óptica, Consejo Superior de Investigaciones Científicas, Serrano 121, 28006 Madrid, Spain

**Abstract.** Non-mydratiac retinal imaging is an important tool for diagnosis and progression assessment of ophthalmic diseases. Because it does not require pharmacological dilation of the patient's pupil, it is essential for screening programs performed by non-medical personnel. A typical camera is equipped with a manual focusing mechanism to compensate for the refractive errors in the eye. However, manual focusing is error prone, especially when performed by inexperienced photographers. In this work, we propose a new and robust focus measure based on a calculation of image anisotropy which, in turn, is evaluated from the directional variance of the normalized discrete cosine transform. Simulation and experimental results demonstrate the effectiveness of the proposed focus measure.

© 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: 10.1117/1.JBO.17.7.076021]

Keywords: focus measure; eye fundus; retinal image; anisotropy; discrete cosine transform.

Paper 12148 received Mar. 2, 2012; revised manuscript received Jun. 19, 2012; accepted for publication Jun. 19, 2012; published online Jul. 17, 2012.

## 1 Introduction

Ocular fundus imaging has long played a key role in the documentation, diagnosis, and progression assessment of ophthalmic diseases. With the advent of digital cameras, ophthalmic imaging changed dramatically. Among the advantages of digital imaging are the ease and speed of access to data, fast and exact duplication, archiving and transmission, and digital image analysis techniques. Altogether, these advantages set the foundations for modern ophthalmology in the framework of telemedicine.

Fundus cameras can be mydratiac or non-mydratiac. Mydratiac fundus cameras require pharmacological dilation, while non-mydratiac cameras use a near-infra-red (NIR) viewing system to exploit the patient's natural dilation in a dark room.<sup>1</sup> Infra-red light is used to preview the retina on a video monitor. Once the monitor's image is focused and aligned, a flash of visible light from a Xenon arc lamp is fired and the image is captured. Non-mydratiac fundus cameras are equipped with a focusing mechanism that displaces a compensation lens. It is basically an aspheric objective lens design that, when combined with the optics of the eye, matches the image plane to the eye fundus. The focus control of the fundus camera is used to compensate for refractive errors in the subject's eye. Until recently,<sup>2</sup> these cameras were entirely operated manually with the focusing mechanism assisted by a split-line visual aid. Manual focusing is error prone, especially when performed by inexperienced photographers, and may lead to images that require additional restoration or enhancement.<sup>3</sup> The auto-focus feature offered in new retinal cameras is a significant advance that ultimately leads to a more robust imaging system, especially for medical screening purposes. On the other hand, the auto-focus feature still relies on the split-line mechanism, whereas in this work we

propose a passive focus measure (FM) completely based on image analysis. For further details on fundus imaging, the reader is referred to Refs. 1, 4, and 5.

In Ref. 6 we studied the performance of several state-of-the-art no-reference image quality metrics for eye fundus imaging. The most interesting finding relates to the importance of directional properties with image quality. In other words, the measure of anisotropy as a quality metric. This was proposed by two co-authors of this paper (Gabarda and Cristóbal) in Ref. 7 and it represents an important step forward in the area of no-reference quality metrics. However, given the properties of the NIR fundus focusing system, the FM should be robust to noise (spatial and temporal) and to changes in illumination and contrast. Furthermore, real-time imaging is constrained by the overhead required to compute the directional Rényi entropy as described in Ref. 7. Therefore, in this work we made use of our findings on the directional dependency of eye fundus images against defocus to define a new and robust FM based on the directional variance of the normalized discrete cosine transform (DCT). The FM proposed here could impact the design of portable retinal cameras with autofocus function or the manufacturing of low-cost retinal cameras because they would require fewer optical components.

### 1.1 Focusing

In a single-lens optical imaging system operating within the paraxial regime, the process of focusing consists in adjusting the relative position of the object, the lens, the image sensor, or a certain combination of the three to obtain a focused image. Let  $f(x, y)$  be the focused image of a planar object and  $g_i(x, y)$  a sequence of images recorded for a sequence of camera parameter settings. The eye fundus is actually a curved surface; however, in our case  $f(x, y)$  corresponds to a small region of the fundus so it can be considered as an isoplanatic

Address all correspondence to: Andrés G. Marrugo, Universitat Politècnica de Catalunya, Group of Applied Optics and Image Processing, Department of Optics and Optometry, Violinista Vellsolà 37, 08222 Terrassa, Spain. Tel: +34937398678; E-mail: andres.marrugo@upc.edu

patch.<sup>8</sup> We consider the variation of only one camera parameter at a time—either the lens position or the focal length. The acquired set of images can be expressed by convolution

$$g_i(x, y) = (f * h_i)(x, y), \quad i = 1, \dots, m, \quad (1)$$

where  $h_i(x, y)$  is the point spread function (PSF) of the blur in the  $i$ th observation. In a practical imaging system, the image magnification and mean image brightness may change while focusing even if nothing has changed in the scene. Normalization with respect to these two parameters can be carried out. However, illumination normalization is more easily performed. Image magnification may be neglected because in most practical applications the magnification is less than 3%.<sup>9</sup> Ideally, the best possible case occurs when  $h_i(x, y) = \delta(x, y)$ , therefore  $g_i(x, y) = f(x, y)$ . In practice, all  $h_i(x, y)$  have an unknown low-pass filter effect.

An FM may be understood as a functional defined on the image space which reflects the amount of blurring introduced by  $h_i(x, y)$ . Let  $S$  be the FM with which we look for the “best” image by maximizing/minimizing  $S(g_i)$  over  $i = 1, \dots, m$ . A reasonable FM should be monotonic with respect to blur and robust to noise. Groen et al.<sup>10</sup> used eight different criteria for the evaluation of focus functions. Ideally, the focus function should be unimodal, but in practice it can present various local maxima that can affect the convergence of the auto-focus procedure. Moreover, the focus curve should ideally be sharp at the top and long tailed, which can accelerate the convergence of the screening procedure.

## 2 Related Works

Various FMs have been reported in the literature.<sup>2,9,11–13</sup> They mainly consist of a focus-measuring operator that estimates the sharpness of the image. The image that yields a maximum FM is considered as the focused one. Almost all FMs depend directly on the amount of high-frequency information in the image. The high-frequency components correspond to edge information. On the other hand, their accuracy can deviate depending on the content of the processed images. Because well-focused images have sharper edges, they are expected to have higher-frequency content than blurred ones.<sup>13</sup> The common FMs are based on norm of gradient or second derivative of the image, gray level variance, and Laplacian energy. Surprisingly, little is known about the performance of these methods for fundus imaging, and the literature on this subject is scarce.

To the best of our knowledge, there exist only two published works that deal with autofocus in retinal imaging;<sup>2,14</sup> however, they use conventional mydiatic imaging in the visible spectrum, which is not our case. In Ref. 14 they do not propose a FM; instead they use several preprocessing operations to improve the performance of traditional FMs for segmentation purposes. On the other hand, in the recent work by Moscaritolo et al.,<sup>2</sup> they propose a filtering technique to assess sharpness of optic nerve head images; however, they do not compare with other methods. In this section we briefly summarize five notable approaches—including that of Moscaritolo et al.<sup>2</sup>—for later comparison with our proposed method.

The first FM  $S_1$  was proposed in Ref. 2. It may be defined mathematically as

$$S_1 = \text{Var} \left( z_{\text{med}} |g_i - z_{\text{lp}}(g_i)| \right), \quad (2)$$

where  $z_{\text{lp}}$  is a low-pass filtering of  $g_i(x, y)$ ,  $z_{\text{med}}$  is a nonlinear median filter of the absolute value  $|\cdot|$  of the difference for removing noise, and  $\text{Var}(\cdot)$  is the variance. Another important measure is the  $l_2$ -norm of image gradient, also called the energy of gradient, defined as

$$S_2 = \sum_x \sum_y \left[ \frac{\partial g_i(x, y)}{\partial x} \right]^2 + \left[ \frac{\partial g_i(x, y)}{\partial y} \right]^2. \quad (3)$$

The third measure is the Laplacian energy. It can analyze high frequencies associated with image edges and is calculated as

$$S_3 = \sum_x \sum_y [\nabla^2 g_i(x, y)]^2. \quad (4)$$

Nayar and Nakagawa<sup>15</sup> proposed a noise-insensitive FM based on the summed modified Laplacian operators. When two second partial derivatives with respect to horizontal and vertical directions have different signs, one offsets the other and the evaluated focus value is incorrect. The method is a modification to obtain the absolute value of each second partial derivative as

$$S_4 = \sum_x \sum_y \left[ \left| \partial^2 \frac{g_i(x, y)}{\partial x^2} \right| + \left| \partial^2 \frac{g_i(x, y)}{\partial y^2} \right| \right]. \quad (5)$$

The frequency-selective weighted median (FSWM) filter<sup>16</sup> is a high-pass nonlinear filter based on the difference of medians. It is well known as a nonlinear edge detector that removes impulsive noise effectively. The FSWM uses several nonlinear subfilters having a weight according to the frequency acting like a bandpass filter as

$$z_F(x) = \sum_j^P \beta_j \hat{z}_j(x), \quad (6)$$

where  $z_F(x)$  is the FSWM filter,  $P$  is the number of subfilters,  $\beta_j \in \mathbb{R}$ , and  $\hat{z}_j(x)$  is the weighted median filter. The FM is produced by summing FSWM results,  $F_x$  and  $F_y$ , applied to an image along the horizontal and vertical directions as

$$S_5 = \sum_x \sum_y (F_x^2 + F_y^2). \quad (7)$$

Subbarao and Tyan<sup>17</sup> analyzed the robustness of three FMs: the image variance (not included here),  $S_2$ , and  $S_3$ . They recommended using  $S_3$  because of its tolerance to additive noise; however, the differences among individual measures were not significant. There are many other FMs, such as the wavelet-based FM proposed in Ref. 12 or the mid-frequency discrete cosine FM in Ref. 11, but they were not included in our study either because of their lack of robustness to noise or their complex implementation. For a review and evaluation of FMs in natural images, the reader is referred to Refs. 9 and 13.

## 3 Discrete Cosine Transform

DCT is an invertible, linear transformation  $\mathcal{T}: \mathbb{R}^N \rightarrow \mathbb{R}^N$ . An image is transformed to its spectral representations by projection onto a set of orthogonal two-dimensional (2-D) basis functions. The amplitudes of these projections are called the DCT

coefficients. Let  $g(x, y)$ , for  $x = 0, 1, 2, \dots, M - 1$  and  $y = 0, 1, 2, \dots, N - 1$ , denote an  $M \times N$  image and its DCT denoted by  $\mathcal{T}[g(x, y)]: G(u, v)$ , given by the equation

$$G(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} g(x, y) \alpha(u) \alpha(v) \cos \left[ \frac{(2x+1)u\pi}{2M} \right] \cos \left[ \frac{(2y+1)v\pi}{2N} \right], \quad (8)$$

where

$$\alpha(\xi; A) = \begin{cases} \sqrt{\frac{1}{A}} & \xi = 0, \\ \sqrt{\frac{2}{A}} & \text{otherwise,} \end{cases} \quad (9)$$

where  $A = \{M, N\}$  depending on variables  $u$  and  $v$ , respectively. Low-order basis functions represent low spatial frequencies, while those of higher orders represent the higher spatial frequencies (Fig. 1). Therefore, low-order coefficients depict slow spatial variations in image intensity, while those of higher orders depict rapid variations.

The DCT is closely related to the discrete Fourier transform (DFT), a standard tool in signal processing, and has been reported as a suitable transform for spectral-based focusing algorithms.<sup>18</sup> However, the DCT has a greater energy compaction property than the DFT, i.e., most of the image information tends to be concentrated in a few low-frequency DCT coefficients. This is also why the JPEG compression standard is based on the DCT. In addition, many efficient schemes for the computation of DCT exist,<sup>19</sup> and hardware implementations are commonly available.<sup>20</sup>

### 3.1 Normalized DCT

The normalized DCT<sup>21</sup> of an image is defined as

$$\tilde{G}(u, v) = \tilde{\mathcal{T}}[g](u, v) = \frac{|\mathcal{T}[g](u, v)|}{\sum_{(u,v)} |\mathcal{T}[g](u, v)|}. \quad (10)$$

This normalization is important because it leads to invariance to changes in the contrast of the image. This can be shown with the following: let  $g'(x, y) = cg(x, y)$ , where  $c$  is a non-zero scaling factor. Given that the DCT is linear, the normalized DCT of  $g'$  is

$$\tilde{\mathcal{T}}[g'](u, v) = \frac{c|\mathcal{T}[g](u, v)|}{c\sum_{(u,v)} |\mathcal{T}[g](u, v)|} = \tilde{\mathcal{T}}[g](u, v), \quad (11)$$

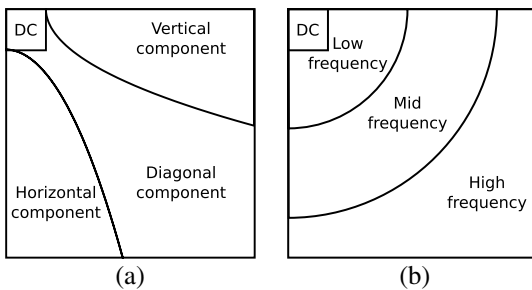


Fig. 1 Relationship between DCT coefficients and frequency components of an image.

which implies that the normalized DCT is contrast invariant and any measure based on this transform as well.

For illustrating the nature of blurring and the behavior of the DCT, we take the red channel from a sharp RGB fundus image (because it resembles more the NIR image) and simulate the imaging system as a linear shift-invariant system to acquire a sequence of images by varying the lens position. This was carried out by means of Fresnel propagation. In Fig. 2, we show the original sharp image, image patches of both the sharp and blurred images, and their DCT spectra (in the same log scale). Notice how the spectrum changes—there is less high- and mid-frequency content in the blurred image spectrum. In addition, in the original spectrum there are some favored orientations in the mid- and low-frequency coefficients, while in the blurred spectrum they seem to become more uniformly distributed. Another important feature is that in the blurred spectrum the coefficients related to high frequency have decreased significantly, and, as described in Sec. 2, many FMs are actually based on the idea of emphasizing high frequencies. While this may be true in theory, in practice there will always be noise contributing to the high-frequency content due to different acquisition conditions. Furthermore,

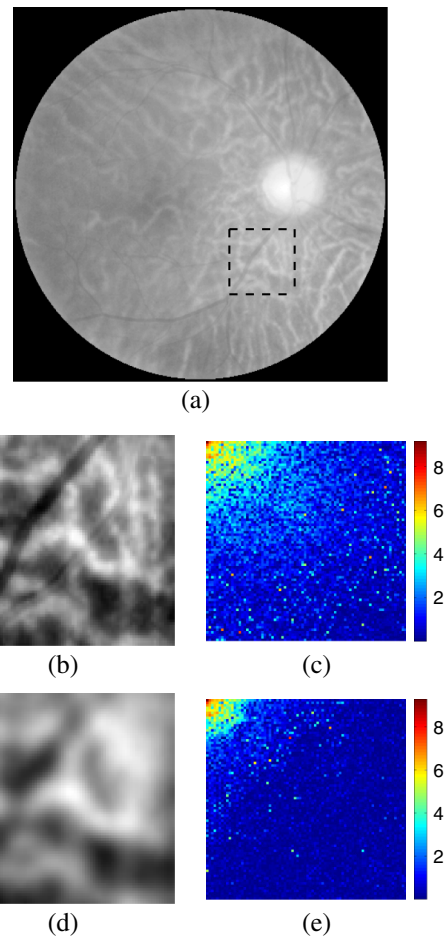


Fig. 2 (a), Original sharp fundus image (R channel from RGB fundus image). (b), ROI from sharp image, and (c), its DCT spectrum. (d), ROI from blurred image, and (e), its DCT spectrum. For visualization purposes both spectra are shown in log scale. Coefficients with higher values are shown in red and those with lower values are shown in blue. The blurred image spectrum is dominated by low-order coefficients.

given that the focusing mechanism involves acquiring a sequence of images, there will be spatial and temporal variations of noise.

## 4 Focus Measure

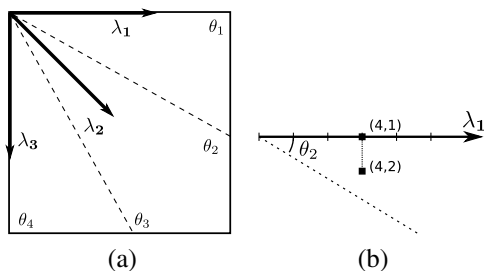
### 4.1 Measure of Anisotropy

As we have seen in the previous example, the overall nature of blurring can be described as a low-pass filtering that tends to break down the characteristic anisotropy of the original image. The FM proposed here aims to quantify this anisotropy dependence based on the normalized DCT of the image.

To define our measure, we introduce some notation. From Eq. (10),  $\tilde{G}(u, v)$  is the normalized DCT of  $g(x, y)$  of size  $N \times N$ , and  $\lambda_j$ , for  $j = 1, 2, 3$ , is a vector along one of the three main orientations of the spectrum depicted in Fig. 3. We will restrict our study to angular partitions of the spectrum roughly equivalent to vertical, diagonal, and horizontal components of the image space. Our measure of anisotropy mainly consists in calculating a difference of weighted coefficients along these orientations. Let  $\tilde{G}_j = \{\tilde{G}(u, v) : \theta = \arctan(\frac{v}{u}), \theta_j \leq \theta < \theta_{j+1}, j = 1, 2, 3\}$  be the set of DCT coefficients located between  $\theta_j$  and  $\theta_{j+1}$  angles, for  $\theta_j \in \{0 \text{ deg}, 30 \text{ deg}, 60 \text{ deg}, 90 \text{ deg}\}$ . The function  $\psi_{\lambda_j}(\cdot)$  takes as input  $\tilde{G}_j$ , performs orthogonal projection of all its elements along vector  $\lambda_j$ , and averages the elements that after projection fall on the same discrete  $(u, v)$  coordinates. With  $\psi_{\lambda_j}(\cdot)$  we seek to compact the information around the three main orientations in a one-dimensional vector of  $N$  elements. To illustrate, let us compute  $\psi_{\lambda_1}(\tilde{G}_1) = [\psi_{\lambda_1}^1, \psi_{\lambda_1}^2, \dots, \psi_{\lambda_1}^N]^T$ , where  $\tilde{G}_1$  is the set of coefficients located between  $\theta_1 = 0 \text{ deg}$  and  $\theta_2 = 30 \text{ deg}$ . In Fig. 3(b), we show the projection of the coefficient with coordinates (4,2) along  $\lambda_1$ . After projection, this coefficient has coordinates (4,1). Therefore, the element  $\psi_{\lambda_1}^4 = \text{mean}[\tilde{G}(4, 1), \tilde{G}(4, 2)]$ . Consequently, we can stack all  $\psi_{\lambda_j}$  to form the following matrix,

$$\Psi = \begin{bmatrix} \psi_{\lambda_1}^1 & \psi_{\lambda_2}^1 & \psi_{\lambda_3}^1 \\ \psi_{\lambda_1}^2 & \psi_{\lambda_2}^2 & \psi_{\lambda_3}^2 \\ \vdots & \vdots & \vdots \\ \psi_{\lambda_1}^N & \psi_{\lambda_2}^N & \psi_{\lambda_3}^N \end{bmatrix}.$$

Note that the first element of each vector corresponds to the dc coefficient. This coefficient does not convey any directional information of the image; however, we decided to keep it in the matrix for the sake of completeness. To obtain a measure



**Fig. 3** (a) Vectors along the main directions of the DCT. (b) Projection of a coefficient along  $\lambda_1$ .

of anisotropy—the FM itself—from  $\Psi$  we compute the variance of the weighted sum of the columns, computed as the matrix product  $\mathbf{w}\Psi$ ,

$$S_a(g) = \text{Var}(\mathbf{w}\Psi) = E[(\mathbf{w}\Psi - \mu)^2], \quad (12)$$

where  $\mathbf{w} = [w_1, w_2, \dots, w_N]$ ,  $E$  is the expected value, and  $\mu$  is the mean of the matrix product  $\mathbf{w}\Psi$ . The vector  $\mathbf{w}$  can be regarded as a weighting procedure and with it we aim to achieve robustness to noise and illumination variation.

### 4.2 DCT Coefficient Weighting

The first issue to address is the selection of a suitable  $\mathbf{w}$ . In DCT-based pattern recognition, robustness is achieved by means of coefficient truncation.<sup>22</sup> It is known that low frequencies are related to illumination variation and smooth regions, and high frequencies represent noise as well as small variations (like edge and details) of the image. The middle-frequency coefficients contain useful information of basic structure; therefore these are suitable candidates for recognition.<sup>23</sup> Consequently, a trade-off between low-frequency and high-frequency truncation should be achieved to obtain a robust FM that is monotonic with respect to blur, unimodal, and at the same time robust to noise and illumination variations.

We decided to find a  $\mathbf{w}$  that meets our requirements based on a training set of  $m$  images. This can be formulated as an optimization problem. The goal would be to find the vector  $\mathbf{w} = [w_1, w_2, \dots, w_N]$  that simultaneously optimizes  $K$  objective values  $\{J_1(\mathbf{w}), J_2(\mathbf{w}), \dots, J_K(\mathbf{w})\}$ . Every objective value  $J_k(\mathbf{w})$  is formulated so that the FM  $S_a$  decreases with respect to blur,  $S_a(g_i^k) > S_a(g_{i+1}^k) \forall i = 1, \dots, m$ . There are  $K$  subsets of  $g_i(x, y)$  all generated in the same way as described in Eq. (1), but they differ in that every  $k$  stands for a different kind of noise degradation, except for  $k = 1$ , the noise-free case. In other words, we want to find a  $\mathbf{w}$  that guarantees monotonicity of  $S_a$  with respect to blur under different types of noise. The objective values are implicitly defined in terms of permutations of the ordered set  $H = \{S_a(g_1), S_a(g_2), \dots, S_a(g_m)\}$ . Thus, the reference permutation is  $\pi_r = \{1, 2, \dots, m\}$ , and any other arbitrary permutation of  $H$  violates the decreasing property of  $S_a$  with respect to blur. As a result, our goal is to find a  $\mathbf{w}$  that produces permutations  $\pi_k$  for all  $K$  types of noise equal to that of  $\pi_r$ . The objective value is defined as the  $l_1$ -norm of the difference between  $\pi_r$  and  $\pi_k$ ,

$$J_k(\mathbf{w}) : \sum_j^m |\pi_r(j) - \pi_k(j)|. \quad (13)$$

It is zero for two identical permutations, and approaches zero as  $\pi_k$  approaches  $\pi_r$ . This is the same for all  $J_k(\mathbf{w})$ ; hence our single aggregate objective function<sup>24</sup> is the weighted linear sum of all  $J_k(\mathbf{w})$ , where all weights are equal to 1.

The solution to this problem is not a straightforward task, as the search space is multivariate and a unique global optimum cannot be guaranteed to exist. Therefore, we solved it using a probabilistic metaheuristic approach called simulated annealing.<sup>25</sup> It provides an acceptably good solution in a fixed amount of time. Each step of the algorithm replaces the current solution by a random nearby solution, chosen with a probability that depends both on the difference between the corresponding function values and on a global parameter  $T$  (called the temperature),

that is gradually decreased during the process. The dependency is such that the current solution changes almost randomly when  $T$  is large, but increasingly downhill as  $T$  goes to zero. (For further details see Ref. 24.)

### 4.3 Implementation

Typically, FMs are applied to a region called the focusing window, which is much smaller than the image. To achieve real-time computation, we decided to implement our measure by dividing the focusing window into subwindows. The measure is computed in the following manner:

1. The focusing window is divided into non-overlapping sub-images of size  $16 \times 16$ . This is chosen so that the most basic structures of the image fit in the subwindows.
2. Each subwindow image is transformed with the normalized DCT, and the FM  $S_a$  is computed.
3. An overall FM  $\bar{S}_a$  is computed by taking the mean of all  $S_a$  values from the subwindows.

According to this implementation, the parameter  $\mathbf{w}$  consists of 16 elements. The considered noise degradations for the procedure described in Sec. 4.2 are Gaussian noise ( $\sigma^2 = 0.001$ ), speckle noise ( $\sigma^2 = 0.001$ ), and impulsive noise ( $d = 0.01$ ). The resulting  $\mathbf{w}$  is shown in Fig. 4. As expected, the first two coefficients are practically zero. Observe the distribution of  $\mathbf{w}$  instead of the individual values per coefficient. This means that a strong emphasis should be put to mid-frequency coefficients. It is perhaps not surprising that the distribution resembles a bandpass filter. This finding is consistent with the work in Ref. 9, where they showed that bandpass filtering causes the FMs to have sharp peaks while retaining monotonicity and unimodality. Interestingly, these weights also resemble the band pass response of the contrast sensitivity function of the human visual system. In the DCT domain, different approaches have been considered for computing visually optimized coefficients for a given image.<sup>26</sup> A major feature of our approach is the fast computation of the FM. The average execution time per frame, in MATLAB implementation on a PC with a 2.66-GHz Intel Core 2 Duo processor, is 40 ms. In most cases this is sufficient; however, if needed, implementation in a low-level programming language could significantly reduce the execution time. In addition, because we divide the focusing window into subwindows, our implementation could be further improved by taking advantage of large parallel architectures such as in graphics processor unit computing.

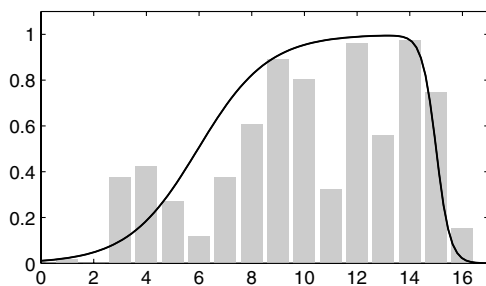


Fig. 4 DCT coefficient weights obtained from the optimization procedure. The distribution resembles a bandpass filter.

## 5 Results

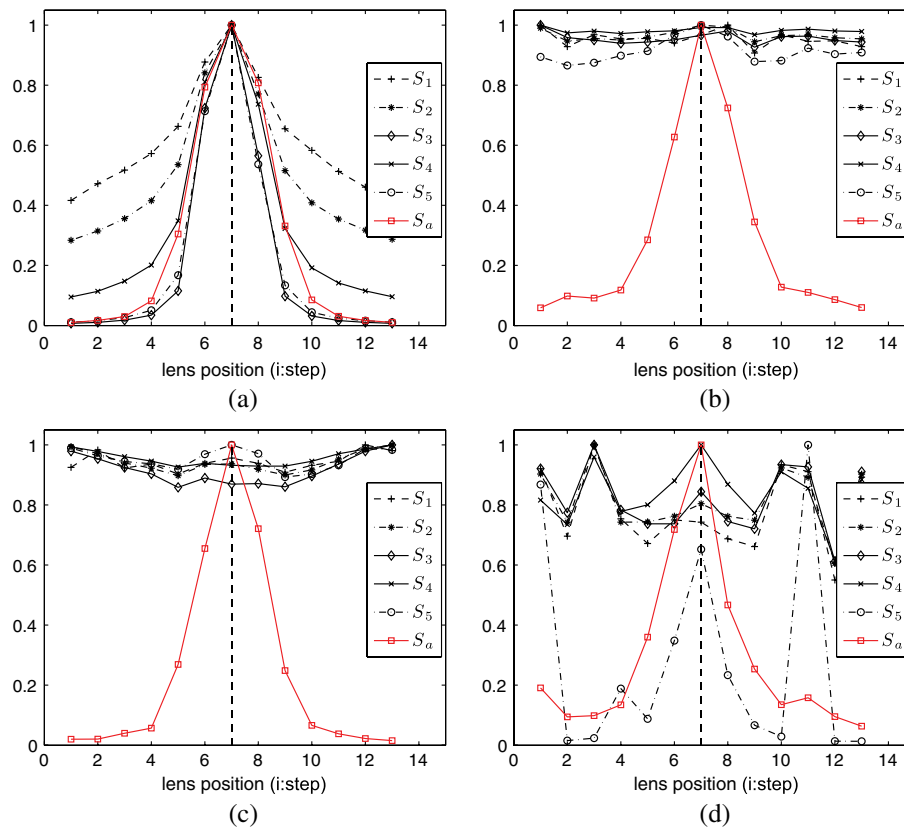
### 5.1 Simulated Images and Robustness Assessment

To evaluate the robustness of our proposed FM  $S_a$  we simulated the focusing procedure. We generated a sequence  $g_i(x, y)$  for  $i = 1, \dots, m$  from the red channel of a sharp RGB fundus image and propagated it at different distances through a linear imaging system of fixed focal length by means of Fresnel propagation. This is equivalent to displacing the lens or the sensor to look for the optimal focus position. From this noise-free sequence, we generated six additional sequences by corrupting it with two levels of three different types of noise: Gaussian, speckle, and impulse noise. We carried out this procedure for 20 retinal images for a total of 140 focusing sequences. Ideally, a noise-robust FM should produce the same (or similar) focusing curve for both the noise-free and corrupted sequences. To quantify the similarity between two focusing curves  $S_r$  and  $S_c$  we used the zero-lag normalized cross-correlation defined as

$$R(S_r, S_c) = \frac{\sum_i S_r(i) \cdot S_c(i)}{\sqrt{\sum_i S_r^2(i) \cdot \sum_i S_c^2(i)}}, \quad (14)$$

where  $r$  stands for the reference curve computed from the noise-free sequence and  $c$  the curve computed from the noise-corrupted sequence. The output is 1 in the case of perfect correlation and 0 for no correlation at all. The reason for the zero-lag calculation, as opposed to the regular cross-correlation by sliding dot product, is that we need the maxima of the curves to coincide in the horizontal position as well as the matching of the profiles.

All FMs were computed using a focusing window of  $128 \times 128$  pixels located over retinal structures. In Fig. 5, we show an example to illustrate the robustness assessment of the FMs. The FM curves represent the normalized measure value over the search space for different lens positions. The highest value should be obtained when the lens is in the optimal focus position identified by the dashed vertical line. As the lens gets farther from the optimal position, the measure value should decrease proportionally to the distance. It comes as no surprise that all measures performed sufficiently well in the noise-free sequence shown in Fig. 5(a), where all curves follow a typical bell shape with a unique maximum. However, in the curves shown in Fig. 5(b)–5(d) where the focusing sequence is corrupted by different types of noise, the proposed FM  $S_a$  clearly outperforms the other measures in terms of monotonicity and unimodality. Notice that under Gaussian and speckle noise [Fig. 5(b)–5(c)], the  $S_a$  curves are nearly identical to the noise-free  $S_a$  curve in Fig. 5(a). Without jumping to conclusions, this result is interesting because it graphically shows the robustness of the proposed FM. The results for all of the 140 sequences are summarized in Table 1. Each value represents the average cross-correlation obtained for all 20 sequences corrupted with a specified type and level of noise for a particular FM. The overall average for each FM is shown in the last column. These results provide further evidence that the proposed FM  $S_a$  has a considerable robustness to noise, with an overall performance value of 0.929 and an exceptional 0.996 for the sequence corrupted with Gaussian noise with  $\sigma^2 = 0.001$ . The second and third best FMs were  $S_4$  and  $S_1$ , with overall values of 0.781 and 0.502, respectively. In comparison with  $S_a$  these values represent a moderate



**Fig. 5** Focus measures curves for the simulated images. The dashed vertical line indicates the correct focused position. (a) Noise-free images. (b) Gaussian noise ( $\sigma^2 = 0.001$ ). (c) Speckle noise ( $\sigma^2 = 0.001$ ). (d) Impulse noise ( $d = 0.01$ ).

**Table 1** Average normalized cross-correlation results for noise robustness assessment of focus measures from 140 sequences generated from 20 retinal images corrupted with different types and levels of noise. The three best FMs are in bold type.

	Gaussian ( $\sigma^2 = 0.001$ )	Gaussian ( $\sigma^2 = 0.005$ )	Speckle ( $\sigma^2 = 0.001$ )	Speckle ( $\sigma^2 = 0.005$ )	Impulse ( $d = 0.01$ )	Impulse ( $d = 0.05$ )	Overall average
$S_1$	0.554	0.486	0.635	0.422	0.477	0.438	<b>0.502</b>
$S_2$	0.524	0.499	0.468	0.408	0.476	0.462	0.473
$S_3$	0.449	0.444	0.370	0.359	0.420	0.417	0.410
$S_4$	0.784	0.782	0.750	0.746	0.836	0.791	<b>0.781</b>
$S_5$	0.495	0.380	0.495	0.304	0.795	0.362	0.472
$S_a$	0.996	0.939	0.997	0.992	0.979	0.667	<b>0.929</b>

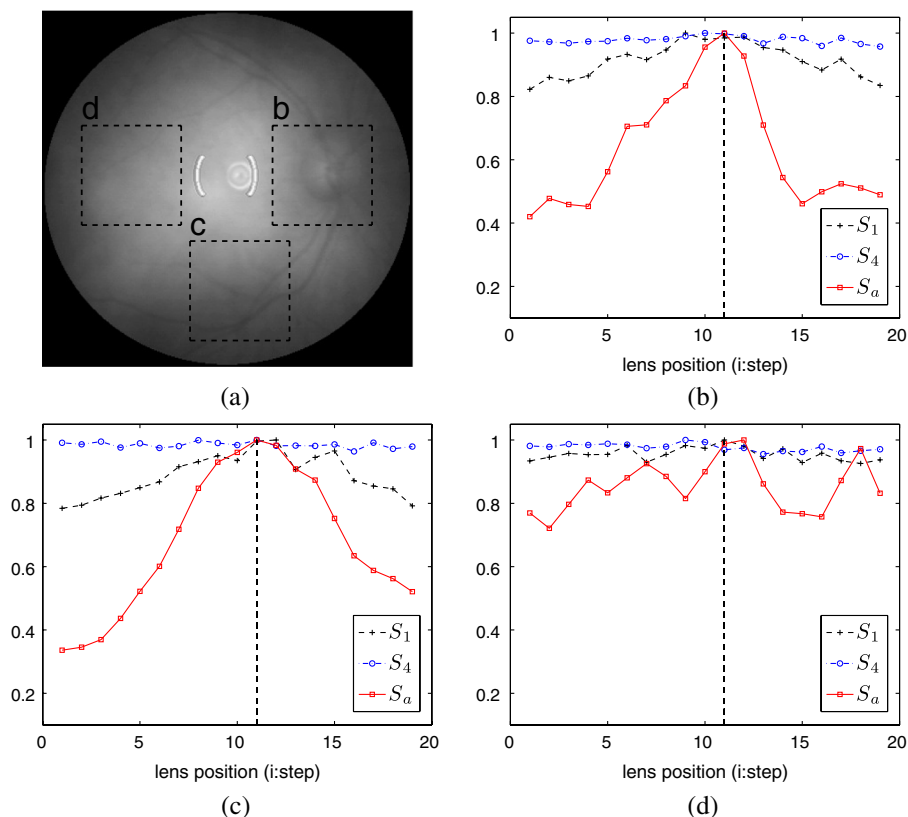
to mild noise robustness. In the following section, we use these two FMs to compare with  $S_a$  in real images.

## 5.2 Real Images

In this subsection we show the results obtained from real NIR-focusing eye fundus images. The images have a relatively low signal to noise ratio (SNR) which justifies the need for a robust FM. All images were acquired using a digital fundus camera system (TRC-NW6S, Topcon, Tokyo, Japan) taking the video output from the infrared focusing system with a resolution of  $640 \times 480$ . The focusing system enables a compensation

range of  $-13D:12D$  in normal operation. For strong myopia or hyperopia, two additional compensation lenses are available to compensate the ranges:  $-12D: -33D$  and  $+9D: +40D$ , respectively. The image sequences analyzed here were acquired by means of an in-house assembled motor mechanism for the displacement of the compensation lens across the whole range for normal operation.

It is well known that as a person ages the crystalline lens of the eye gradually gets opacified, obstructing the passage of light. This is called a cataract. A complete loss of transparency is observed only in advanced stages in untreated patients. In early stages of cataracts retinal examination is considered



**Fig. 6** Focus measure curves obtained by placing the focusing window over different regions of (a) the retinal image. The dashed vertical line indicates the correct focused position. Areas (b) and (c) are located over prominent retinal structures, whereas (d) is located over a relatively uniform region.

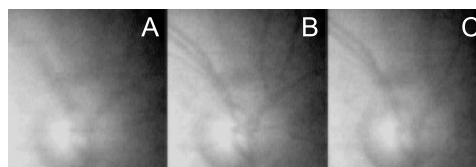
practicable—however, it is not without difficulty. For this reason, we decided to test our focusing method on healthy young subjects and elderly subjects with first signs of cataracts, not only to demonstrate its applicability on real images, but to assess its limitations as well. In this work we show results from five representative subjects of ages 27, 40, 68, 70, and 81 years for a total number of 10 eye fundi.

First we show the effects of placing the focusing window on different regions of the retinal image. A retinal image has distinct sharp structures such as the blood vessels and the optic disk, as opposed to the relatively uniform background. No FM is reliable without placing the focusing window on top of structures with edges, a fact easily appreciable from the three focusing curves shown in Fig. 6, which were computed from the right eye fundus of the 27-year-old subject. The optimal focus position identified by the dashed vertical line was verified via the split-line focusing mechanism. The  $S_a$  curves computed from regions (b) and (c) are clearly reliable in terms of monotonicity and unimodality and coincide on the optimal focus position. Conversely, the  $S_1$  and  $S_4$  curves fail to produce a reliable profile against the  $S_a$  curves that display a steeper peak at the optimal focus position, evidence of the measure's robustness to noise. In contrast, all measures computed from region (d) are unusable because they are mainly given by noise.

To illustrate the link between the focusing curves and the image quality, in Fig. 7 we show three image details depicting the optic disk region for three different focusing positions. The image detail in Fig. 7(b) corresponds to the focused image (optimal focus position 11 in the  $S_a$  curve Fig. 6). Notice how this image is properly focused: it has sharp details such as the blood vessels. The other two images are blurred,

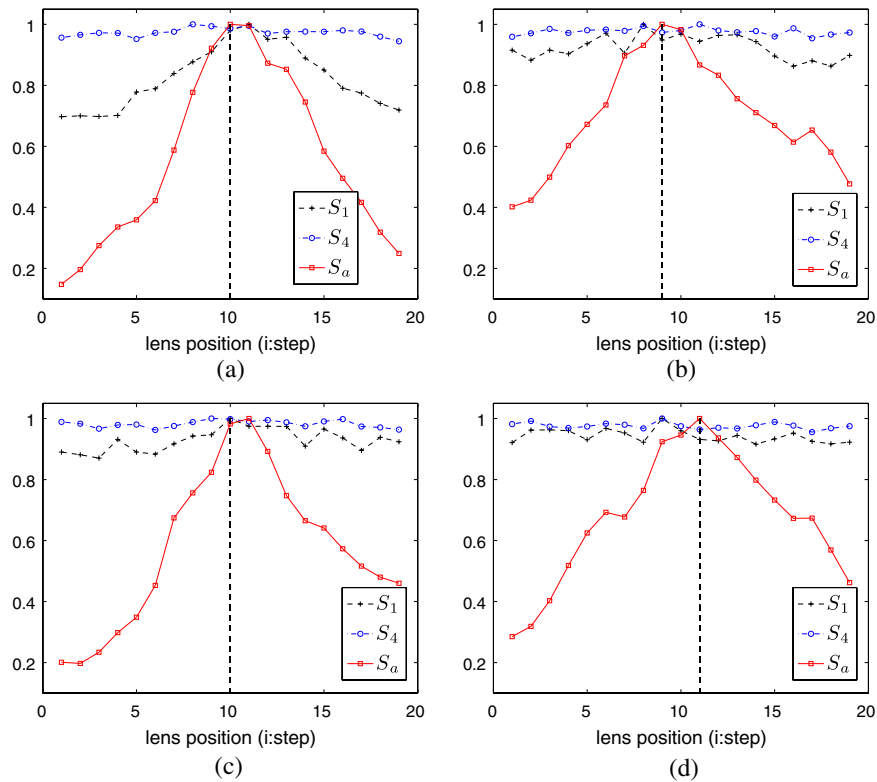
demonstrating the consistency of the  $S_a$  curves with image quality or sharpness. The result that emerges from this example is that to effectively locate the best-focused image, homogeneous regions should be avoided. An adaptive technique, based in an edge detector for example, could prove useful for detecting such prominent structures and therefore candidate regions for applying the focusing technique automatically. The focusing curves shown hereafter, however, were all computed from a focusing window located manually over retinal structures.

To further analyze the performance of the FM in Fig. 8, we show the focusing curves obtained from four of the five subjects; the ages are shown in the figure caption. From the four cases shown only in one [Fig. 8(c)], the  $S_a$  measure peak did not coincide precisely with the optimal focus position. However, the error is no more than a single position. The FMs curves of  $S_1$  and  $S_4$  are generally flatter than those of  $S_a$  which in a focus search strategy is not wanted because of the difficulty to properly distinguish the optimum position in a coarse or initial search. From the curves in Fig. 8, we can also note that there appears to be little difference between the curves from young

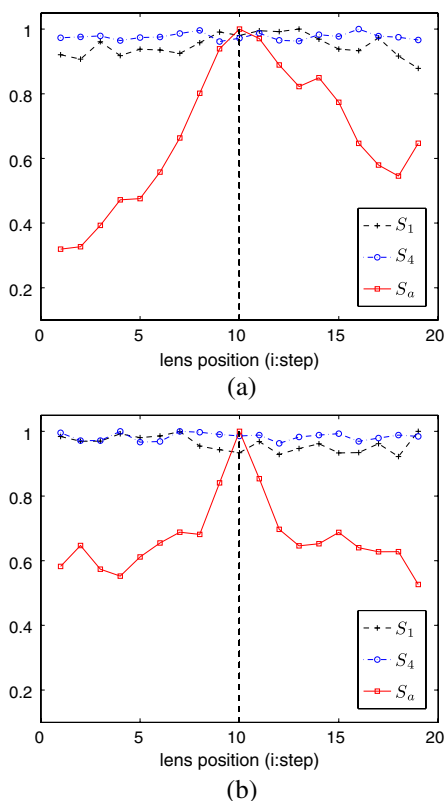


**Fig. 7** Image detail from Fig. 6 for different focusing positions. (a) 6, (b) 11 (optimal focus), and (c), 15. The positions are in reference to Fig. 6(b)–6(c).





**Fig. 8** Focusing curves obtained from four subjects with ages 27 (a) 40 (b) 68 (c), and 70 (d) years. The dashed vertical line indicates the correct focused position.



**Fig. 9** Focusing curves obtained from the 81-year-old subject for each eye fundus. In the right eye (a), the crystalline lens has been extracted and replaced with an intraocular lens. The left eye (b), is in an early stage of cataract. The dashed vertical line indicates the correct focused position.

and elderly subjects. In Fig. 9, we show the focusing curves obtained from the 81-year-old subject for both eye fundi. This case is interesting on its own because in the right eye [Fig. 9(a)], the crystalline lens has been extracted and replaced with an intraocular lens, whereas the left eye [Fig. 9(b)], is in an early stage of cataract. While both focusing curves are able to successfully identify the optimal focus position, the curve in Fig. 9(b) is certainly flatter throughout most of the search space. This is most likely due to the difference in visibility and clarity from both eyes. In general, from the comparison against  $S_1$  and  $S_4$  it can clearly be stated that the proposed FM  $S_a$  outperforms them in the considered cases.

A close examination of the results reveal that the shape of the focusing curve is not exclusively given by the degree of defocus, but by the state of the subject's eye and the analyzed region of the fundus as well. This is important because it conditions the strategy for searching the optimal focus position. Finally, even though the results seem to indicate that the FM could be successfully applied to both young and elderly subjects, further research on a higher number and variety of subjects is necessary. Additionally, we report here that we encountered some difficulty in the procedure with the elderly subjects related to sustaining fixation during the acquisition procedure. From an initial number of six subjects one was excluded from all calculations due to this condition. Patient inability to successfully establish fixation is a true challenge in fundus photography, and dealing with it is out of the scope of this work.

## 6 Conclusion

In this paper, a new robust FM for nonmydiatic retinal imaging has been proposed. It is based on a measure of anisotropy, mainly the weighted directional variance of the normalized

DCT. The weights were calculated by means of an optimization procedure to maximize the noise robustness of the FM. Not only were the resulting weights in agreement with previous works,<sup>9</sup> but they also provide a key insight into the design of noise-invariant FMs. By both simulation and real fundus imaging, we demonstrated the robustness and the accuracy of the proposed FM, which clearly outperformed the other considered measures. The findings presented here may have a number of implications for the design and operation of auto-focusing in modern retinal cameras. Finally, in this study we included several young and elderly subjects to assess the limitations of the proposed FM. Even though we found no significant differences between the focusing curves, there was some difficulty in the acquisition of images from the elderly mainly given by inability to sustain fixation. As with all such studies, there are limitations that offer opportunities for further research. Adapting our method to these variations within the patient population is a goal worth pursuing.

### Acknowledgments

This research has been partly funded by the Spanish Ministerio de Economía y Competitividad and Fondos FEDER (project DPI2009-08879) and projects TEC2010-09834-E and TEC2010-20307. The first author also thanks the Spanish Ministerio de Educación for an FPU doctoral scholarship. The authors especially thank the Pérez-Cabré and the Sisquella-Cabré families for their cooperation in the experimental session.

### References

1. T. J. Bennett and C. J. Barry, "Ophthalmic imaging today: an ophthalmic photographer's viewpoint—a review," *Clin. Exp. Ophthalmol.* **37**(1), 2–13 (2009).
2. M. Moscaritolo et al., "An image based auto-focusing algorithm for digital fundus photography," *IEEE Trans. Med. Imag.* **28**(11), 1703–1707 (2009).
3. A. G. Marrugo et al., "Retinal image restoration by means of blind deconvolution," *J. Biomed. Opt.* **16**(11), 116016 (2011).
4. R. Bernardes, P. Serranho, and C. Lobo, "Digital ocular fundus imaging: a review," *Ophthalmologica* **226**(4), 161–181 (2011).
5. M. D. Abramoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.* **3**, 169–208 (2010).
6. A. G. Marrugo et al., "No-reference quality metrics for eye fundus imaging," in *CAIP 2011, Lect. Notes Comput. Sci.* **6854**, 486–493 (2011).
7. S. Gabarda and G. Cristóbal, "Blind image quality assessment through anisotropy," *J. Opt. Soc. Am. A* **24**(12), B42–B51 (2007).
8. P. Bedggood et al., "Characteristics of the human isoplanatic patch and implications for adaptive optics retinal imaging," *J. Biomed. Opt.* **13**(2), 024008 (2008).
9. M. Subbarao, T. S. Choi, and A. Nikzad, "Focusing techniques," *Opt. Eng.* **32**(11), 2824–2836 (1993).
10. F. C. A. Groen, I. T. Young, and G. Lighthart, "A comparison of different focus functions for use in autofocus algorithms," *Cytometry* **6**(2), 81–91 (1985).
11. S. Y. Lee et al., "Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning," *IEEE Trans. Circuits System. Video Technol.* **18**(9), 1237–1246 (2008).
12. J. Kautsky et al., "A new wavelet-based measure of image focus," *Pattern Recogn. Lett.* **23**(14), 1785–1794 (2002).
13. V. Aslantas and R. Kurban, "A comparison of criterion functions for fusion of multi-focus noisy images," *Opt. Commun.* **282**(16), 3231–3242 (2009).
14. P. Liatsis and P. Kantartzis, "Real-time colour segmentation and autofocus in retinal images," in *ELMAR, 47th International Symposium*, pp. 13–18, IEEE, Zadar, Croatia (2005).
15. S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(8), 824–831 (1994).
16. K. S. Choi, J. S. Lee, and S. J. Ko, "New autofocusing technique using the frequency selective weighted median filter for video cameras," *IEEE Trans. Consum. Electron.* **45**(3), 820–827 (1999).
17. M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 864–870 (1998).
18. N. N. K. Chern, P. A. Neow, and M. H. Ang, "Practical issues in pixel-based autofocusing for machine vision," in *IEEE Int. Conf. Robotics Automation*, Vol. **3**, pp. 2791–2796, IEEE, Singapore (2001).
19. G. K. Wallace, "The jpeg still picture compression standard," *IEEE Trans. Consum. Electron.* **38**(1), 18–34 (1992).
20. J. Ramirez et al., "A new architecture to compute the discrete cosine transform using the quadratic residue number system," in *IEEE Int. Symp. on Circuits and Systems*, Vol. **5**, pp. 321–324, IEEE, Geneva, Switzerland (2000).
21. M. Kristan et al., "A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform," *Pattern Recogn. Lett.* **27**(13), 1431–1439 (2006).
22. Z. Lian and M. J. Er, "Illumination normalisation for face recognition in transformed domain," *Electron. Lett.* **46**(15), 1060–1061 (2010).
23. W. Chen, M. J. Er, and S. Wu, "Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain," *IEEE Trans. Syst. Man Cybern. B Cybern.* **36**(2), 458–466 (2006).
24. A. Suppaitnarm et al., "A simulated annealing algorithm for multiobjective optimization," *Eng. Optimiz.* **33**(1), 59–85 (2000).
25. V. Granville, M. Krivanek, and J.-P. Rasson, "Simulated annealing: a proof of convergence," *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(6), 652–656 (1994).
26. A. B. Watson, "Perceptual optimization of DCT color quantization matrices," in *ICIP94, IEEE Int. Conf. Image Proc.*, Vol. **1**, pp. 100–104, IEEE, Austin, TX (1994).