amcs

PARAMETRIC LOGARITHMIC TYPE IMAGE PROCESSING FOR CONTRAST BASED AUTO–FOCUS IN EXTREME LIGHTING CONDITIONS

CORNELIU FLOREA, LAURA FLOREA

Image Processing and Analysis Laboratory
University Politehnica of Bucharest, Splaiul Independenţei 313, Bucharest, Romania
e-mail: `{corneliu.florea,laura.florea}@upb.ro`

While most of state-of-the-art image processing techniques were built under the so-called classical linear image processing, an alternative that presents superior behavior for specific applications comes in the form of Logarithmic Type Image Processing (LTIP). This refers to mathematical models constructed for the representation and processing of gray tones images. In this paper we describe a general mathematical framework that allows extensions of these models by various means while preserving their mathematical properties. We propose a parametric extension of LTIP models and discuss its similarities with the human visual system. The usability of the proposed extension model is verified for an application of contrast based auto-focus in extreme lighting conditions. The closing property of the named models facilitates superior behavior when compared with state-of-the-art methods.

Keywords: logarithmic image processing, digital camera, auto-focus.

1. Introduction

Logarithmic Image Processing (LIP) models are part of a larger category of non-linear image processing techniques and were introduced (Jourlin and Pinoli, 1987) in opposition to image processing with real based operations. While the later combination, named Classical Linear Image Processing (CLIP) (Lim, 1990), proves its limitations under specific circumstances, like upper range overflow, LIP models—due to their mathematical properties—are capable of dealing with such cases. Subsequently, many extensions and alternative models with numerous practical applications appeared. In order to allow more flexibility, parametric extensions of LIP models were constructed (Panetta *et al*., 2011; Deng, 2012). While the initial, Jourlin–Pinoli model specifically followed a logarithmic one and placed itself under the homomorphic theory elaborated by Oppenheim (1965), the more recent parametric extensions require *additional* and *restrictive* constraints in order to keep within the same theory.

In this paper we propose an extension of LIP models by parametrization. Four main contributions are claimed. First we will revisit the algebraic substrate of logarithmic models and we will formulate a set of simple conditions that allow extensions of LIP models while preserving the cone space structure (i.e., keeping within homomorphic theory). Secondly, using the said conditions, we will elaborate a parametric extension of the LIP models. Thirdly, we will discuss the similarity of the proposed model with the human visual system. Fourthly, to demonstrate the practical usability of the proposed model, we will integrate it in an application of contrast-based auto-focus for extreme lighting conditions which proves to be superior to the state of the art.

The remainder of the paper is organized as follows. In Section 2, an overview of the existing LIP models is provided, followed by a discussion on mathematical background. In Section 3 we will discuss the mechanism for generating a parametric extension for these models while keeping them within homomorphic theory. Section 4 describes the proposed method for focusing in extreme lighting conditions and in Section 5 we discuss the achieved results. Section 6 summarizes the proposed solutions and discusses further continuation paths.

2. Logarithmic image processing models

The first LIP model was constructed by Jourlin and Pinoli (1987; 1988) starting from the equation of light passing through transmitting filters. The model was further used in various applications, like background removing

(Wu and Jeng, 2002), image enhancement (contrast and sharpness improvement) (Jourlin and Pinoli, 1995; Deng *et al.*, 1995), and so on. An extensive review of advances and applications for LIP may be found in the work of Deng *et al.* (1995) as well as Pinoli and Debayle (2007). Summarizing, the strong points of LIP models that lead to successful applications include

1. strong mathematical properties, namely, affiliation to a *closed algebraic system*;

2. *resemblance with the Weber–Fechner perception law.*

In Section 3.1, we will show that LTIP models have stronger properties.

Later advances in LIP witness Deng's interpretation of the LIP model from an entropy point of view (Deng, 2009) and a proof of similarity with the Giga-vision image sensor (Deng, 2012).

In parallel, other derivations of these models appeared. First, the cone-space structure was completed to a vector space structure (Pătraşcu and Voicu, 2000), while more recently the generative logarithmic function was replaced by a logarithmic-like one with the benefits of a simplified calculus (Vertan *et al.*, 2008). This extension preserves the same behavior, yet it differs in the nature of the generative function. Thus, in the remainder of the paper we will collectively name the models Logarithmic Type Image Processing (LTIP) ones.

Meanwhile, parametric extensions of LTIP models have been proposed. The first significant result in parametrization of LTIP models may be considered the one reported by Panetta *et al.* (2008; 2011). They proposed a parametric extension of the Jourlin model, named PLIP (Parametric LIP). Compared with the corresponding Jourlin initial LIP model operations, the upper range value of the grey-levels, D, has been replaced by a parametric linear function $\gamma(D) = A + B(D)$. The introduction of the parameters offers greater flexibility for defining new operations but has the cost of loosing the cone space structure. The model, if different from the initial one (i.e., $\gamma(D) \neq \Lambda(D)$), is no longer an extension of homomorphic systems introduced by Oppenheim (1965), which are the basis for LTIP functionality.

Another parametrization was introduced by Deng (2012). He extends the Jourlin model in the framework of homomorphic theory (Oppenheim, 1965), which he calls GLIP (Generalized LIP) based on the similarity with the model of the Giga-vision sensor. Also this extension requires an additional constraint (see Table 2) over the upper limit to preserve the cone space structure. Thus, we note that none of the previously mentioned parametric extensions presents a natural way of preserving the closed algebraic system (i.e., the cone space).

2.1. Vector/cone space structure. Taking into account the necessity for additional constraints required by state-of-the-art parametric LIP models, we shall start by revisiting the algebraic background of LTIP models.

The mathematical origin of an LTIP model lies in homomorphic theory developed by Oppenheim. Such a model may be constructed by redefinition of the operational laws, the addition and the scalar multiplication, or by means of a function named a generative one. In this subsection we will present a set of results that define the conditions which guarantee achieving a vector(cone) space structure when applied to the named generative function.

Let us consider a function, $\Phi : \mathcal{D}_\phi \to E$. Within this choice, the set $\mathcal{D}_\phi$ is the one-dimensional image definition set. For color (multi-dimensional) images, the discussion may refer to each plane independently. Typically, if the image values have intensity meaning, the set is bounded (e.g., $[0, 255]$ or $[0, 1]$).

The function Φ defines the model structure and maps the image definition set (also named the tone or gray set), $\mathcal{D}_\phi$, onto a subset of real numbers, E. Following the theory elaborated by Oppenheim (1965; 1967), the two basic operations (addition of two elements of the set, $\oplus$, and multiplication, $\otimes$, with an outer, typically real scalar α) are defined over the given set, $\mathcal{D}_\phi$, as follows:

$$\Phi(u \oplus v) = \Phi(u) + \Phi(v), \tag{1}$$

$$\Phi(\alpha \otimes u) = \alpha\Phi(u), \tag{2}$$

where $u, v \in \mathcal{D}_\phi$ and $\alpha \in K \subseteq \mathbb{R}$.

Equations (1) and (2) are the conditions that must be fulfilled by a homomorphism between two similar algebraic structures. The simplest solution is to consider the function Φ as a bijection and, hence, to have the laws uniquely determined. Any logarithmic or logarithmic like model following these equations was named generalized LIP (Deng, 2012).

With respect to the bijectivity constraint (thus, the existence of Φ^{-1}), the definition laws are determined by

$$u \oplus v = \Phi^{-1}\left(\Phi(u) + \Phi(v)\right), \tag{3}$$

$$\alpha \otimes u = \Phi^{-1}\left(\alpha\Phi(u)\right). \tag{4}$$

Of practical importance for LTIP models is the closing property of both addition and scalar multiplication. This states that the sum of any two images should lead to another valid image and, respectively, any amplified or attenuated image should be an image:

$$\begin{array}{ll} \forall u, v \in \mathcal{D}_\phi, & z = u \oplus v \Rightarrow z \in \mathcal{D}_\phi, \\ \forall u \in \mathcal{D}_\phi, \forall \alpha \in K, & z = \alpha \otimes u \Rightarrow z \in \mathcal{D}_\phi. \end{array} \tag{5}$$

Given the two operative laws, $\oplus$, $\otimes$, the vector set $\mathcal{D}_\phi$ and the outer scalar set K, the formal definition of

Table 1. Some of the existing logarithmic type image processing models with basic operations. D is the upper bound of the image definition set (typically $D = 255$ for `unsigned int` representation or $D = 1$ for `float` image representation).

Model	Domain	Isomorphism	Addition, $u \oplus v$	Scalar multiplication, $\alpha \otimes u$
Jourlin and Pinoli (1988)	$\mathcal{D}_\phi = (-\infty; D]$	$\Phi_J(x) = -D \log \frac{D}{D-x}$	$u + v + \frac{uv}{D}$	$D - D\left(1 - \frac{u}{D}\right)^\alpha$
Pătraşcu and Voicu (2000)	$\mathcal{D}_\phi = (-1; 1)$	$\Phi_P(x) = \frac{1}{2} \log \frac{1+x}{1-x}$	$\frac{u+v}{1+uv}$	$\frac{(1+u)^\alpha - (1-u)^\alpha}{(1+u)^\alpha + (1-u)^\alpha}$
Vertan *et al.* (2008)	$\mathcal{D}_\phi = [0; 1)$	$\Phi_V(x) = \frac{x}{1-x}$	$1 - \frac{(1-u)(1-v)}{1-uv}$	$\frac{\alpha u}{1+(\alpha-1)u}$

Table 2. Parametric LTIP models. Note that the Panetta model has a cone space structure only if $\Lambda(D) = \gamma(D)$, while Deng enforces $\Phi(x + y, D) = \Phi(x, D_x) + \Phi(y, D_y)$ to achieve the same structure. Parameter variation of the Deng model "shifts" the original one, for the Panetta model the main effect is that of the scaling (as it changes the upper limit of the image definition set), while for the proposed parametrization the changes are in curvature strength.

Model	Isomorphism	Addition	Scalar multiplication
Panetta *et al.* (2008)	$\Phi(x) = -\Lambda(D)\left(\log\left(1 - \frac{x}{\Lambda(D)}\right)\right)^\beta$	$u + v + \frac{uv}{\gamma(D)}$	$\gamma(D) - \gamma(D)\left(1 - \frac{u}{\gamma(D)}\right)^\alpha$
Deng (2012)	$\Phi(x, D_x) = -D \log\left(1 - \frac{x}{D_x}\right)$	$D\left(\frac{u}{D_u} + \frac{v}{D_v} + \frac{uv}{D_u D_v}\right)$	$D\left(1 - \left(1 - \frac{u}{D_u}\right)^\alpha\right)$
Proposed	$\Phi_m(x) = \frac{x^m}{1-x^m}$	$\sqrt[m]{1 - \frac{(1-u^m)(1-v^m)}{1-u^m v^m}}$	$u \exp\left(\frac{1}{m} \log \frac{\alpha}{1+(\alpha-1)u^m}\right)$

the vector space implies several properties (see Hefferon, 2008, Section II.1). More precisely, the vector addition has to be associative and commutative and should have a neutral element and inverse element, while the scalar multiplication should be distributive with respect to vector addition in the field of vectors and in the field of scalars, should respect field multiplication and have an identity element. These properties do hold under the bijectivity constraint.

The existence of the addition identity element, u_0, implies further conditioning over the mapping function, Φ. The mentioned restriction is a consequence of the isomorphic behavior:

$$\forall u \in \mathcal{D}_\phi, \ \exists u_0, \ u \oplus u_0 = u \Leftrightarrow \Phi(u_0) = 0. \quad (6)$$

The existence of an addition inverse element, u^- is conditioned by a symmetry towards 0 of the generative function. This property makes a difference between vector and cone space. However, since this is not of paramount importance for practical applications, in many cases the LTIP model has a cone structure.

In much the same way as for addition, the identity element of the scalar multiplication has to be 1:

$$\exists \alpha_1, \forall u \in \mathcal{D}_\phi - \{u_0\}, \ \alpha_1 \otimes u = u \Leftrightarrow \alpha_1 = 1. \quad (7)$$

It can be determined that the set of *sufficient conditions* that needs to be fulfilled by a generative function so as to produce a usable logarithmic-type image processing model (and to preserve the named properties, which are of practical importance) is as follows:

- Φ should be bijective;

- the target E should be at least $[0, \infty)$ in the case of a cone structure;

- $\Phi(u_0) = 0$.

Any new model that completes under the given rule will produce a cone space structure.

For models that have only a cone space structure, the subtraction is defined as follows:

$$u \ominus v = \Phi^{-1}\left(\Phi(u) - \Phi(v)\right), \quad (8)$$

where u is enforced to be larger than v, $u > v$.

The models mostly used in image processing (Jourlin and Pinoli, 1988; Pătraşcu and Voicu, 2000; Vertan *et al.*, 2008) are summarized in Table 1.

Given the found mathematical boundaries of the generative function, there are many ways to derive parametric models. In the next subsection, we will present a modality for constructing specific parametric extensions.

3. Parametrization

While Panetta *et al.* (2008) aim simply at more flexibility and Deng (2012) uses the Giga-vision model for parametrization, we construct our proposal from a mathematical point of view. The basic result that allows us to extend LTIP models by parametrization comes from algebraic theory (see Hefferon, 2008, Section III.2) and was introduced by Florea *et al.* (2009). It states that the composition of two valid homomorphisms leads to another homomorphism. Let there be $\psi : \mathcal{D}_\phi{}^1 \to \mathcal{D}_\phi{}^2$, a homomorphism from $\mathcal{D}_\phi{}^1$ to $\mathcal{D}_\phi{}^2$, and let $\phi : \mathcal{D}_\phi{}^2 \to E_2$ be a homomorphism from $\mathcal{D}_\phi{}^2$ to E_2. Then the composite function $\rho : \mathcal{D}_\phi{}^1 \to E_2, \rho = \phi \circ \psi =$

$\phi(\psi)$ is a homomorphism from $\mathcal{D}_\phi{}^1$ to E_2. With such a construction, we choose ϕ to be the generative function, Φ, of a known LTIP model, while ψ may be a real function with a bounded domain and a target set; the result is a valid new LTIP model. One may go even further: it is not really necessary to have the ψ function a homomorphism, but only a bijective function; even in such a case the result is a valid generative LTIP model function.

The parametrization is naturally achieved if the ψ function is a parametric one that in all the cases uses the same sets. Such an example may be the family of "power(gamma)-type" functions:

$$\psi_m : [0, 1) \to [0, 1), \qquad \psi_m(x) = x^m, \qquad (9)$$

$\forall m \in (0, +\infty)$. We note that all members of this family are bijective functions.

By composing this family of functions with the generative function of the logarithmic-like model (Vertan *et al*., 2008), one will obtain a set of parametric LTIP models. The base function is

$$\Phi_m : \quad [0, 1) \to [0, +\infty),$$
$$\Phi_m(x) = \Phi\left(\psi_m(x)\right) = \frac{x^m}{1 - x^m}. \qquad (10)$$

The inverse function is

$$\Phi_m^{-1}(y) = \psi_m^{-1}\left(\Phi^{-1}(x)\right) = \sqrt[m]{\frac{y}{1 + y}}. \qquad (11)$$

The mathematical formulas of the so-generated model are found if one replaces the formulas (10) and (11) in Eqns. (3), (4) or (8). Hence, the general formula for addition, $\oplus_m$, becomes

$$u \oplus_m v = \psi_m^{-1}(\psi_m(u) \oplus \psi_m(v))$$
$$= \sqrt[m]{1 - \frac{(1 - u^m)(1 - v^m)}{1 - u^m v^m}} \qquad (12)$$

where $\oplus$ is the logarithmic-like addition. The detailed form of the subtraction, $\ominus_m$, is

$$u \ominus_m v = \psi_m^{-1}(\psi_m(u) \ominus \psi_m(v))$$
$$= \sqrt[m]{\frac{(u^m - v^m)}{1 + u^m v^m - 2v^m}}, \quad u > v \qquad (13)$$

Some illustrative graphical examples of the subtraction are presented in Fig. 1.

Scalar amplification is done as

$$\alpha \otimes_m u = u \exp\left(\frac{1}{m} \log \frac{\alpha}{1 + (\alpha - 1)u^m}\right). \qquad (14)$$

We have to note that the "power" function is thereby a choice and any other bijective parametric function will do. In the case used, for $m = 1$ the logarithmic-like model is obtained.

3.1. Relation with the human visual system. A strong point in the motivation of the utility of LIP models is their similarity with the human visual system. This was noted by Jourlin and Pinoli (1988) and thoroughly discussed by Pinoli and Debayle (2007). Basically, it was shown that the initial LIP subtraction is consistent with the Weber fraction and the Fechner law of perception.

Yet the Weber–Fechner model is not unanimously accepted. Stevens (1961) strongly challenged this practice and argued for the use of the power-law rules (Stevens and Stevens, 1963). Thus we note that the Stevens model is in-line with the proposed parametric extension. On the other hand, Stevens experiments were also questioned (Macmillan and Creelman, 2005), so it does not seem to be a definite answer in this regard. Therefore we shall focus on unanimously accepted facts.

The general shape of the power-law and of the Weber–Fechner model are similar and this similarity is shared by all the *closed* logarithmic and logarithmic-type models. First, there is a general acceptance that the human eye never saturates (Ferwerda *et al*., 1996). The lack of saturation is valid also for any *closed* mathematical model. Furthermore, there is an agreement (both Stevens and Weber–Fechner) that visual quanta in the domain of low light are significantly larger than in the high light domain. This trait is also characteristic to all LTIP models, thus encouraging the use of LTIP models and the proposed parametric extension.

3.2. Edge detection. To prove the utility of the parametric LTIP model we will describe in Section 4 an application for contrast-based auto-focus. Such an application relies on edge detection operators. Motivated by the resemblance of LTIP models with the human visual system, we claim that the edge detection operator should rely on a logarithmic-type subtraction rather than on the classical one.

Edge detection is an important issue in image processing (Fabijańska, 2012). Here, we shall use the luminance-invariant inter-color distance as suggested by Deng *et al*. (1995) for the implementation of a Laplacian operator L. The proposed operator is a modification of the classical (derivative-type) V4–neighborhood Laplacian operator; the proposed Laplacian is the average difference between the value (gray-scale or on each color plane) at the currently processed location $(i; j)$ and its immediate neighboring values from the image I. Mathematically we can express the proposed operator, at the $(i; j)$ coordinates, as

$$L_m(i; j) = \sum_{(k;l) \in V_4} \frac{1}{4}\left(I(i + k; j + l) \ominus_m I(i; j)\right) \qquad (15)$$

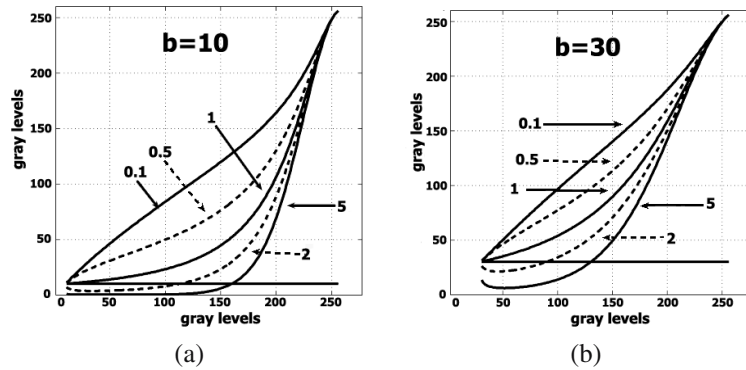assuming that $I(i + k; j + l) > I(i; j)$.

Fig. 1. Graphical representation of $\xi(u, b) = u \ominus_m (u - b)$, with respect to u, for different values of m: $\{0.1, 0.5, 1, 2, 5\}$. In the left panel (a), we used $b = 10$, while in the right (b) $b = 30$. Note that in the higher part of the domain the non-linear subtraction is much more sensitive than the classical subtraction (horizontal solid line), while in the lower part only supra-unitary parameter values lead to less sensitivity.

The implied subtraction is given by Eqn. (13), while the addition (implicitly written in the sum) is the one shown in Eqn. (12). Now, let us note that Eqn. (15) can be re-written as

$$L_m(i; j) = \sum_{(k;l) \in V_4} \frac{1}{4} \xi \left(I_{\min}^{(i,j,k,l)}, \delta_{(i,j)}^{(k,l)} \right), \quad (16)$$

with

$$I_{\min}^{(i,j,k,l)} = \min \left(I(i,j), I(i+k, j+l) \right),$$
$$\delta_{(i,j)}^{(k,l)} = |I(i,j) - I(i+k, j+l)|, \quad (17)$$
$$\xi(u, u - b) = u \ominus_m (u - b).$$

Examples of $\xi(\cdot, \cdot)$ are presented in Fig. 1.

4. Auto-focus

Digital cameras, in their various forms (still cameras, video cameras, mobile phones), have gained significant popularity in the recent years. Nowadays, the main direction seems to be that of decreasing the size and weight of the imaging devices so as to embed them in various other consumer electronics devices. The trend of miniaturization imposes design modifications such as reducing the size of optics and of the photo-sensible area, which leads to more pressure on the image processing pipeline. It should be noted that, due to small camera dimensions and to avoid blur due to shaky hands, mobile phone cameras are usually forced to acquire underexposed images/videos and to amplify them later.

The sharpness of the acquired image is given by the physical properties of the sensor and the optical system and, most importantly, by the auto-focus solution (Ramanath *et al.*, 2005). A camera de-focus causes damage on the acquired images of an observed sample: out-of-focus parts of the sample appear blurred, whereas in-focus parts look relatively sharp.

4.1. Contrast based auto-focus. Currently, two main categories of auto-focus algorithms exist in cameras: passive and active. Active systems measure the distance to the subject independently of the optical system, and, subsequently, adjust the optical system for correct focus. Passive focusing systems determine the correct lens position by analyzing the images entering through the optical system. Two major passive solutions exist: phase detection and contrast measurement. The algorithm proposed goes within the line of the last category, namely, contrast measurement.

Motivated by the similarity with the human visual system, this paper proposes a reinterpretation of a classical focus measurement under the parametric LTIP framework. A similar method was proposed by Fernandes *et al.* (2010), who used Jourlin's model for a shape-from-focus solution for normally exposed microscopic images. Instead, we make use of the previously described parametric extension to provide more adaptability of the algorithm. We differentiate furthermore by the fact that we use the superior sensitivity of LTIP models adaptively on strong and weak luminance and in order not to enhance noise.

The contrast measurement is achieved by measuring sharpness within a sensor field, through the lens. The intensity difference between adjacent pixels of the sensor naturally increases with a correct image focus. For different lens positions, different values for contrast are computed. The subject-to-camera distance is determined for the sharpest recorded image, via optical system settings. This estimation is known as the depth from focus and a review of the known techniques may be found in the works of Krotkov (1987), Subbarao and Tyan (1998), as well as Sun *et al.* (2005).

The sharpness (or the local contrast) of an image is measured with so-called Focus Measure Function (FMF). The FMF encodes the relative proximity of the system to

the focus position with respect to various lens positions. In practice, the FMF may be either of a derivative type (and then we look for the maximum) or in the family of correlation methods (Vollath, 1987), the case when the minimum is looked for. Once the measure is available, the FMF optimum with respect to the lens position must be searched for. A review of the problem specific techniques to determine the maximum of the FMF is given by Svahn (1996). We note that the FMF resolution (i.e., one/two pixels) should be consistent with the image real resolution.

Between possible FMFs, there is the sum of a modified Laplacian, which has been initially introduced to estimate the shape of the subject (Nayar and Nakagawa, 1994). The sum of modified Laplacian, $\Upsilon(I_s)$, within a convex subset of the image $I_s \subset I$ is the sum of the absolute values of the second grey-scale derivative, computed with a discrete approximation of the Laplacian, $L(i; j)$ (which is the second derivative in the horizontal and vertical directions of an image):

$$\Upsilon(I_s) = \sum_{(i;j) \in I_s} L(i; j). \tag{18}$$

Florea and Florea (2011) proposed a solution for focusing on bright objects, hence overexposed images. In the current work, we extend the algorithm to underexposed images, too. The exposure case is known at the beginning of the acquisition. If we know the solution for overexposure, in the case of the underexposure we choose to simply use the negative of the image ($u \to u' = D - u$).

In the simplest form, the computation of the Laplacian $L(i; j)$ is given by Eqn. (15), thus forming a parametric function with respect to m. The global FMF used is therefore parametric: $\Upsilon(I_s) \to \Upsilon_m(I_s)$.

The simple use of the Laplacian, as defined by Eqn. (15) and in the formulation of Fernandes *et al.* (2010), will not do much good to the focusing problem. The closing property, written in Eqn. (5) says that the sum of any two numbers between 0 and D will also lead to a number smaller than D. If $D = 255$, this means, for example, that the sum of 235 and 240 will give 248. Then, if we subtract, using Eqn. (13), 240 from 248, we will get 235. This result is beneficial for bright scenes with pale edges, but it also may have the consequence that we are measuring the noise of the image. Hence, we modified the computation of the Laplacian, from $L_m(i; j)$ into $L_m^{cut}(i; j)$, as follows:

$$L_m^{cut}(i; j) = \begin{cases} L_m(i; j), & L_m(i; j) \geq \frac{1}{4} \otimes_m \overline{I_s}, \\ \\ 0, & L_m(i; j) < \frac{1}{4} \otimes_m \overline{I_s}. \end{cases} \tag{19}$$

where $\overline{I_s}$ is the average of the pixels from the I_s patch computed according to the classical image processing model. In this way, the contribution of noise is greatly

reduced, while the LTIP edge detector preserves its special sensibility.

Concluding, the procedure to determine the lens position that provides the maximum sharpness is the following:

- Using the relative exposure of the image, consider the initial or the negative version of the image.

- Initialize the parameter m to the smallest possible value.

- Given an image (or an image patch) I_s with a known lens position, k_0, compute the average $\overline{I_s}$.

- Compute $\Upsilon_m^{k_0}(I_s)$ at the lens position k_0 using Eqns. (18), (19) and (15).

- Determine the parameter value, m, that provides the maximum value $\Upsilon_m(I_s) = \Upsilon_{m_{\max}}(I_s)$. Store the found value, $m_{\max} = m$.

- Using $m_{\max}$, compute $\Upsilon_{m_{\max}}^{k_i}(I_s)$ for various lens positions, k_i, to determine the focus. The focus will be given by the position of the lens, k_i, which maximizes $\Upsilon_{m_{\max}}^{k_i}$.

Regarding the best value, $m_{\max}$, of the parameter, as one may see in Fig. 2, the function, $\Upsilon_m(I_s)$, is uni-modal and in a majority of cases it has the maximum at the first or the last value. We note that the maximum is not at $m = 1$, thus arguing for the use of the parametric extension. Among many parameter estimation methods existing in the literature (see the work of Byrski and Byrski (2012) and the references therein), we found that a simple implementation of the hill climbing algorithm (always go in the sense of the positive slope (Russell and Norvig, 2003)) solves the problem.

Now let us give an insight of into the method. The basic statement is that parametrization allows more adaptability to the scene, hence leading to improved results when compared with the simple re-writing of the classical FMF under logarithmic-like model. The motivation for the procedure lies in the fact that if the scene is static, then the histogram of the blurred image is the blurred version of the sharp image's histogram, hence preserving the positions of the modes (which corresponds to objects). The aim of this adaptation is to get the maximum difference (in the sense of Eqn. (13)) between modes (favoring edges) and to have the minimum difference inside the modes.

Because the proved mathematical model for the out-of-focus blur is the convolution with a Gaussian kernel (Svahn, 1996), the variation of the FMF, including here $\Upsilon(I_s)$, used has the shape of a Gaussian probability density function (i.e., a uni-modal function). An example of the variation is presented in Fig. 3. Again the hill climbing procedure is used for the determination of the maximum of the FMF.
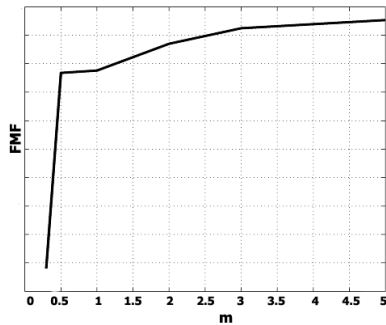
Fig. 2. Variation of the FMF used, $\Upsilon(I_s)$, with respect to parameter m for a static scene. As can be seen, it is expected to achieve better results for $m > 5$, but in that case quantization errors for a discrete space computation become significant and introduce artefacts.
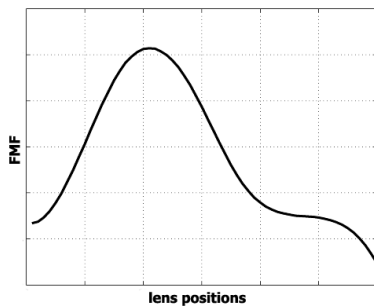

Fig. 3. Variation of the FMF used, $\Upsilon_m(I_s)$, for a fixed value of the parameter, $m = 5$, with respect to the lens position. The sharpest image is attained for the maximum of $\Upsilon_m(I_s)$. The function has been interpolated from the original 12 values to 60 values.

5. Implementation and results

5.1. Databases. We tested the proposed method on two types of databases: one using examples from public databases and another one internally built. While the public databases contain well exposed original images and simulated out-of-focus blur, our database was built by manually adjusting the focus lens, thus containing naturally degraded frames and with unbalanced exposures (over and underexposed).

The public database set contains all useful image sets from the Tampere Image Database (TID2008)[1] (Ponomarenko *et al.*, 2009) and from the CSIQ image database[2] (Larson and Chandler, 2010). We note that all these images have the exposure well balanced and the variation regards the scene type and the content.

TID2008 contains 25 reference images and 1700 distorted images (25 reference images $\times$ 17 types of distortions $\times$ 4 levels of distortions). Among these

[1]http://www.ponomarenko.info/tid2008.htm
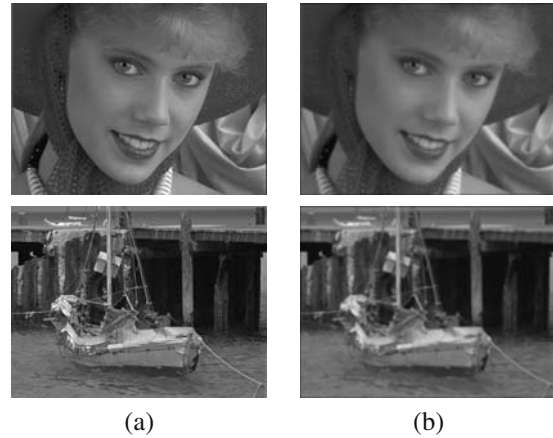[2]http://vision.okstate.edu/?loc=csiq

(a)   (b)

Fig. 4. Example of images from the TDI2008 database: original images (with all objects in focus) (a), Gaussian blurred image simulating objects out of focus (b).

distortions is the Gaussian blur, which was noted as being the model for out-of-focus blur (Svahn, 1996). Therefore, relevant for our work are 25 sets with 5 images each. Examples of images from this database can be seen in Fig. 4.

The CSIQ database consists of 30 original images, each distorted using 6 different types of distortions, including Gaussian blur at 4 to 5 different levels of distortion. Out of this databases resulted 30 sets $\times$ 6 images. Images from this database can be seen in Fig. 5.

For each sequence of images from the public data sets, we built 5 tests by considering 5 different patches of interest (therefore varying the content of the patch).

Since our database concentrates more on unbalanced exposures, it can be considered that it complements the public databases, where all images are well exposed. For acquisition we used a webcam and a professional camera, because both of them permit manual precise lens positioning. We varied the nature of the object aimed to be in focus (average intensity, sharpness) and the distance to the camera. For each such setting, images corresponding to different lens positions were acquired. This made what will be subsequently called a "set of images".

The total database consisted out of 114 sets, 65 with a bright objects and 49 with dark ones. An example of a scene with a bright object may be observed in Fig. 6, while that with a dark one in Fig. 7. The variation in the lens position was done in 11 steps (hence each set consists out of 12 images) for the professional camera and 24 images for the webcam. The images were acquired with reduced shift (scene had mainly the same content).

For each set of images (a sequence of frames of the same scene with different blur), multiple tests were performed, by changing the initial image (assumed different beginning position for the lens). For a sequence of K images, $K - 2$ tests were considered by placing the
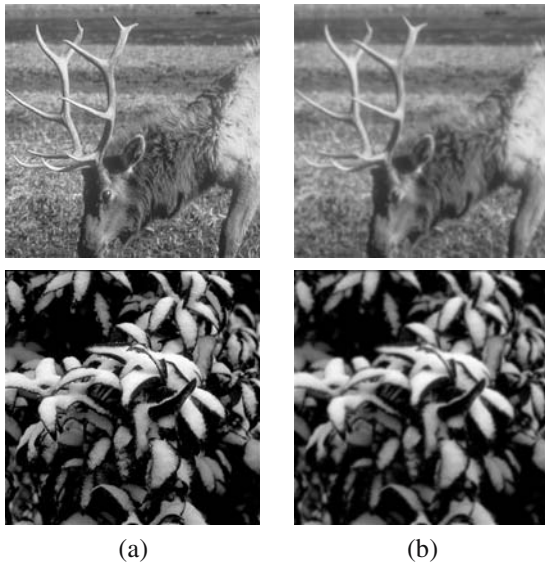
Fig. 5. Example of images from the CSIQ database: original images (with all objects in focus) (a), Gaussian blurred image simulating objects out of focus (b).

Fig. 6. Example of a scene with a bright object and reduced perceived contrast. Top row: object in focus and moving away, middle row: gradient image obtained with the LTIP Laplacian, bottom row: gradient image obtained with the Vollath F4 function. Typically, our method produces a more accurate but flatter FMF, while Vollath—sharper but with fake modes.

initial assumed lens position in all intermediate situations.

5.2. Evaluation procedure. The auto-focus algorithm was implemented on the equivalent gray-level image patch, extracted from the color image. For simplicity, we reduced the range of possible parameter values to $\{0.3, 0.5, 1, 2, 3, 5\}$. This reduction permitted also simplification of the algorithm by the use of look-up-tables for power functions.

The image patches were selected randomly from the scene. In the case of our database, we remove the patches that contain two objects at different depths (hence confusing the algorithm and human observer—a situation encountered in our database). In general, the object of interest is at least 80% of the total patch area. The computation of the average of the image patch $\overline{I_s}$ was performed on a sub-sampled version ($4\times$ smaller) of the image patch. The image patches were chosen as rectangular areas that contained mostly one object.

Various methods to estimate the correctness of an auto-focus algorithm exist. Subbarao and Tyan (1998) propose the so-called Auto-focusing Uncertainty Measure (AUM) and Auto-focusing Root Mean Square (ARMS) error. They mainly encode the sharpness of the FMF function with respect to the lens position and the number of correct focus estimations. Usually, one will get a failure from the auto–focusing algorithm when the FMF function is present, due to various reasons (noise, lack of edges, camera movement) several modes and the hill climbing procedure ends while producing wrong maxima.

For evaluation, we used measures that have a more straight-forward interpretation. We counted in how many

cases the correct position was determined and, in the case of an error, what was the error in the number of steps of the lens position (because an error of one step may get unnoticed by the user).

Testing was performed under specific scenarios, thus containing three categories: the first with cases from public databases (simulated cases), the second with bright objects from our database, and thirdly with dark objects.

For comparison, we considered two other popular FMF measures: a derivative one—the classical sum of modified Laplacian and one from the correlation family, namely, Vollath F4 (Vollath, 1987). These are in general (Sun *et al.*, 2005) perceived as the best performers. More recent solutions, such as those by Kristan *et al.*
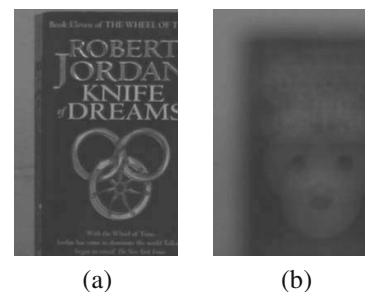

Fig. 7. Example of a scene with a dark object: dark background and bright fonts form a high perceived contrast. Object in focus (a), object out of focus (b).

Table 3. Evaluation performance on the data set made out of images from the CSIQ and TDI2008 databases. The total number of tests is 975. The ground truth is the patch from the original image. The mean error represents the average number, only in the case of a failure, of lens position steps between the reported focus and the true focus position. The total error represents the summation of all errors on all sets.

Method	Correct estimates	Mean error	Total error
Proposed: adaptive m	**93.3%**	**2.6**	**170**
Proposed: fixed $m = 1$	67.7%	2.6	836
Classical Laplace (Nayar and Nakagawa, 1994)	80.3%	2.6	499
Vollath (1987)	88.0%	2.6	304

(2006) or Lee *et al.* (2009), even though they rely on higher complexity they do not manage to overcome older solutions. The search technique used was the same for all FMFs, i.e., the hill climbing procedure.

To determine if the parametric extension is really an improvement, we also included in our comparison the solution that uses $m = 1$. This, approximately, corresponds to the description by Fernandes *et al.* (2010).

5.3. Discussion. The achieved results on the public databases are presented in Table 3, while those for our database can be seen in Tables 4 (bright scene) and 5 (dark scene).

Let us note that one important aspect of the focus, namely focusing speed is missing in the current evaluation because it is related to the convergence technique. Once we decided to use the hill-climbing method (arguments pro and con may be found in the work of Svahn (1996)), the algorithm will go through all lens positions from initialization till convergence regardless of whether the FMF used. Thus, as long as the algorithm converges correctly, it will require the same amount of steps if our function or any other state-of-the- art solution is used.

Regarding the tests on the public database, we note that we outperform all proposed state of the art methods. In this test, the variation was given only by the amount of blur and one can see that, with a very good precision, the proposed method correctly identified the sharpest image. Furthermore, this test showed that the proposed measure, being in an inverse relation with the blur variance, may be a reliable basic metric for an image quality general measure.

While discussing the tests on our database, we stress that the images were acquired with the camera in the auto mode and were not processed afterward, therefore being degraded by typical camera noise (which can be seen in Fig. 6). Also the camera was held in hand. Thus there is a slight content variation in the sequence of images.

Discussing the actual achieved results, we note that the adaptive (parametric) solution significantly out-performed the simple, non-adaptive version that relied on $m = 1$, strongly arguing for the utility of the parametric extension.

Regarding the actual precision, a basic observation is that the FMF for the proposed method is flatter. In association with hill climbing, this behavior has no drawback. However, it has the benefit of being more robust and performs more consistently. As one can see in the mentioned tables, our method outperformed the state-of-the-art solutions on the bright object database. Even when it missed the correct position, it did it with a number of steps. On the darker object data set it was outperformed by the Vollath function in terms of the number of correct detection of the lens position that provides the best focus, but it did better in terms of the error size. This means that, small although Vollath gives fewer errors, they are rather big. Our method leads to more errors but they are smaller that may be less noticeable by the user.

5.4. Algorithm complexity. Since the implementation of the auto-focus algorithm is aimed at embedded systems that can vary in parameters, we evaluate the algorithm performance in operation per pixel and not in seconds. Given an image patch, the Vollath method requires accessing the current location and the next consecutive two ones and two additions and one multiplication. The classical Laplacian requires buffering 3 rows, 4 differences and 3 additions. The proposed method requires also buffering 3 rows and for each of the 4 neighbors a comparison, a difference and incrementing a 2D histogram (which requires a multiplication and 2 additions). Next, to process the 2D histogram each value must pass through 2 LUTs (hence 2 additions). Thus, compared with the classical Laplacian there is an increase of 4 multiplications and 1 addition per pixel and 2 additions and memory for storing the LUTs.

6. Conclusions and future work

In this paper we summarized our contribution to the advance of logarithmic-type image processing models and described their possible extensions with the precise purpose of preserving the mathematical property resulting

Table 4. Correct estimations of the focus position for bright scenes. The total number of tests is 1010. The ground truth resulted after manual annotation. The mean error represents the average number, only in the case of a failure, of lens position steps between the reported focus and the true focus position. The total error represents the summation of all errors on all sets.

Method	Correct estimates	Mean error	Total error
Proposed: adaptive m	**93.1%**	**2**	**139**
Proposed: fixed $m = 1$	87.7%	4.7	584
Classical Laplace (Nayar and Nakagawa, 1994)	84.8%	4.25	652
Vollath (1987)	92.3%	4.5	350

Table 5. Correct estimates of the focus position for dark objects. The total number of tests is 594. The ground truth resulted after manual annotation.

Method	Correct estimates	Mean error	Total error
Proposed: adaptive m	**87.5%**	**2.5**	**185**
Proposed: fixed $m = 1$	83.5%	4.25	416
Classical Laplace (Nayar and Nakagawa, 1994)	85.8%	5.5	464
Vollath (1987)	91.2%	3.66	188

from the definite structure of the vector or cone space. The most important resulting property is that of closing. If ensures special behavior near the edges of the definition set.

From the application point of view, a non-linearity localized especially near the boundary of the value domain makes the approach suitable for processing images that have an important part of the histogram in these ranges; these are images acquired in extreme lighting. We showed the usability of the proposed framework by describing a contrast based auto-focusing algorithm that gave results similar to or superior over state-of-the-art methods.

Regarding further continuation paths, the algorithms are currently in the process of optimization, including adaptation for ASIC (Application Specific Integrated Circuit) acceleration. The latter is rather easy since the framework is pixel-wise oriented, and hence parallelization may be used.

The problem of measuring sharpness has another (more recent) use, namely, multi-sensor (acquisition) image fusion. This topic has many applications such as remote sensing, medical imaging, microscopic imaging and robotics. In a specific embodiment (Li *et al.*, 2010), the criteria for combining information from two or more source images of a scene into a single composite image to extract the focused parts from each multi-focus image and produce one with equal clarity. This is known as multi-focus image fusion and the core of the technique is estimation of the sharpness on image blocks.

Acknowledgment

References

Byrski, W. and Byrski, J. (2012). The role of parameter constraints in EE and OE methods for optimal identification of continuous LTI models, *International Journal of Applied Mathematics and Computer Science* **22**(2): 379–388, DOI: 10.2478/v10006-012-0028-3.

Deng, G. (2009). An entropy interpretation of the logarithmic image processing model with application to contrast enhancement, *IEEE Transactions on Image Processing* **18**(5): 1135–1140.

Deng, G. (2012). A generalized logarithmic image processing model based on the giga-vision sensor model, *IEEE Transactions on Image Processing* **21**(3): 1406–1414.

Deng, G., Cahill, L.W. and Tobin, G.R. (1995). A study of logarithmic image processing model and its application to image enhancement, *IEEE Transactions on Image Processing* **4**(4): 506–512.

Fabijańska, A. (2012). A survey of subpixel edge detection methods for images of heat-emitting metal specimens, *International Journal of Applied Mathematics and Computer Science* **22**(3): 695–710, DOI: 10.2478/v10006-012-0052-3.

Fernandes, M., Gavet, Y. and Pinoli, J.C. (2010). Improving focus measurements using logarithmic image processing, *Journal of Microscopy* **242**(3): 228–241, http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2818.2010.03461.x/abstract.

Ferwerda, J.A., Pattanaik, S.N., Shirley, P. and Greenberg, D.P. (1996). A model of visual adaptation for realistic image synthesis, *SIGGRAPH Conference Proceedings, New Orleans, LA, USA,* pp. 249–258.

Florea, C. and Florea, L. (2011). A parametric non-linear algorithm for contrast based autofocus, *Proceedings of the IEEE International Conference on Intelligent Computer Communication and Processing, ICCP, Cluj, Romania*, pp. 75–82.

Florea, C., Vertan, C., Florea, L. and Sultana, A. (2009). Non-linear parametric derivation of contour detectors for cellular images, *Proceedings of the IEEE International Symposium on Signals, Circuits and Systems, ISSCS, Iaşi, Romania*, Vol. 2, pp. 321–325.

Hefferon, J. (2008). *Linear Algebra*, Web edition, http://joshua.smcvt.edu/math/hefferon.html.

Jourlin, M. and Pinoli, J.C. (1987). Logarithmic image processing, *Acta Stereologica* **6**(1): 651–656.

Jourlin, M. and Pinoli, J.C. (1988). A model for logarithmic image processing, *Journal of Microscopy* **149**(1): 21–35.

Jourlin, M. and Pinoli, J.C. (1995). Image dynamic range enhancement and stabilization in the context of the logarithmic image processing model, *Signal Processing* **41**(2): 225–237.

Kristan, M., Pers, J., Perse, M. and Kovacic, S. (2006). A Bayes-spectral-entropy-based measure of camera focus using a discrete cosine transform, *Pattern Recognition Letters* **27**(13): 1431–1439.

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

Larson, E.C. and Chandler, D.M. (2010). Most apparent distortion: Full-reference image quality assessment and the role of strategy, *Journal of Electronic Imaging* **19**(1): 011006.

Lee, S., Yoo, J., Kumar, Y. and Kim, S. (2009). Reduced energy-ratio measure for robust autofocusing in digital camera, *IEEE Signal Processing Letters* **16**(2): 133–136.

Li, X., He, M. and Roux, M. (2010). Multifocus image fusion based on redundant wavelet transform, *IET Image Processing* **4**(4): 283–293.

Lim, J.S. (1990). *Two Dimensional Signal and Image Processing*, Prentice Hall, Upper Saddle River, NJ.

Macmillan, N. and Creelman, C. (Eds) (2005). *Detection Theory: A User's Guide*, Lawrence Erlbaum, Mahwah, NJ.

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

Oppenheim, A.V. (1965). Superposition in a class of non-linear system, *Technical report*, MIT, Cambridge, MA.

Oppenheim, A.V. (1967). Generalized superposition, *Information and Control* **11**(5,6): 528–536.

Panetta, K., Wharton, E. and Agaian, S. (2008). Human visual system-based image enhancement and logarithmic contrast measure, *IEEE Transactions on Systems, Man, and Cybernetics, B: Cybernetics* **38**(1): 174–188.

Panetta, K., Zhou, Y., Agaian, S. and Wharton, E. (2011). Parameterized logarithmic framework for image enhancement, *IEEE Transactions on Systems, Man, and Cybernetics, B: Cybernetics* **41**(2): 460–472.

Pinoli, J.C. and Debayle, J. (2007). Logarithmic adaptive neighborhood image processing (LANIP): Introduction, connections to human brightness perception, and application issues, *EURASIP Journal on Advances in Signal Processing* **036105**(1), Article ID 36105, DOI: 10.1155/2007/36105.

Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M. and Battisti, F. (2009). A database for evaluation of full-reference visual quality assessment metrics, *Advances of Modern Radioelectronics* **10**(1): 30–45.

Pătraşcu, V. and Voicu, I. (2000). An algebraic model for gray level images, *Proceedings of the Exhibition on Optimization of Electrical and Electronic Equipment, OPTIM, Brasov, Romania*, pp. 809–812.

Ramanath, R., Snyder, W., Yoo, Y. and Drew, M. (2005). Color image processing pipeline: A general survey of digital still camera processing, *IEEE Signal Processing Magazine* **22**(1): 34–43.

Russell, S.J. and Norvig, P. (2003). *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.

Stevens, J. and Stevens, S. (1963). Brightness functions: Effects of adaptation, *Journal of the Optical Society of America* **53**(3): 375–385.

Stevens, S. (1961). To honor Fechner and repeal his law, *Science* **133**(3446): 80–133.

Subbarao, M. and Tyan, J. (1998). Selecting the optimal focus measure for autofocussing and depth-from-focus, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(8): 864–870.

Sun, Y., Duthaler, S. and Nelson, B. (2005). Autofocusing algorithm selection in computer microscopy, *Proceedings of the International Conference on Intelligent Robots and Systems, Edmonton, Canada*, pp. 809–812.

Svahn, F. (1996). *Tools and Methods to Obtain a Passive Autofocus System*, Master's thesis, Technical University of Linkoping, Linkoping, www.viktoria.se/~fresva/documents/master_thesis.pdf.

Vertan, C., Oprea, A., Florea, C. and Florea, L. (2008). A pseudo-logarithmic framework for edge detection, *in* J.B. Talon, S. Bourennane, W. Philips, D. Popescu and P. Scheunders (Eds.), *Advances in Computer Vision*, Lecture Notes in Computer Science, Vol. 5259, Springer-Verlag, Juan-les-Pins, pp. 637–644.

Vollath, D. (1987). Automatic focusing by correlative methods, *Journal of Microscopy* **147**(3): 279–288.

Wu, Q.Z. and Jeng, B.S. (2002). Background subtraction based on logarithmic intensities, *Pattern Recognition Letters* **23**(13): 1529–1536.

Corneliu Florea was born in 1980 in Bucharest. He got his master degree from the University Politehnica of Bucharest in 2004 and the Ph.D. from the same university in 2009. There he lectures on statistical signal and image processing, and runs introductory courses in computational photography and computer vision. His research interests include non-linear image processing algorithms for digital still cameras and computer vision methods for portrait understanding.

Laura Florea received her Ph.D. in 2009 and M.Sc. in 2004 from the University Politehnica of Bucharest. Since 2004 she has been teaching at the same university, where she is currently a lecturer. Her research interests include image processing algorithms for digital still cameras, medical image processing and statistic signal processing theory. She is also a postdoctoral researcher on automatic detection of human emotion by the analysis of the portrait image. Previously she worked on computer aid diagnosis for hip joint replacement.