Autofocusing with the help of orthogonal series transforms

Przemysław Śliwiński
Institute of Computer Engineering, Control and Robotics,
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland

***Abstract*—An autofocus algorithm employing orthogonal series expansions is proposed. Several instances of the generic algorithm, based discrete trigonometric, polynomial and wavelet series, are reviewed. The algorithms are easy to implement in the transform coders used in digital cameras. Formal analysis of the algorithm properties is illustrated in experiments. Some practical issues are also discussed.**

***Keywords*—Autofocusing, discrete orthogonal systems, fast orthogonal transform, transform coding.**

I. Introduction

The advent of CCD/CMOS sensors shifted the digital imaging-related researches from a *niche* of mainly 'academic-oriented problems' to the mainstream topics and resulted in various applications in both applied and physical sciences and in consumer electronic devices. A plethora of the 'off-the-shelf' theoretical results developed in various disciplines like *signal and image processing*, *image analysis*, *harmonic analysis* and *information theory*, *probability theory* and*mathematical statistics*, or eventually in *control theory*, have found, in consequence, their applications in many real devices. We refer the reader to the special issue of *IEEE Signal Processing Magazine*, [1], for the selection of introductory articles encompassing the variety of digital imaging.

In the paper we consider the problem of autofocusing in a generic digital camera. The proper, fast and reliable focusing algorithm is a *conditio sine qua non* of a 'good image' not only from *aesthetic* vantage point, but also in automated *shape from focus* applications, where the three dimensional scene is recovered from the sequences of images, and where the precise information about the depth of scene is of the paramount importance, see *e.g.* [2,3].

The problem of autofocusing attracted many authors and various approaches were proposed. An important class of AF algorithms (to which belongs the algorithm proposed in the paper) are those in which the focus functions is evaluated exclusively from the data acquired by the image sensor (that is, the optical/electric path used to capture an image is also exploited to assess the focus function).

Intuitively, the image (or its fragment, referred further to as the *region of interest*) is 'in-focus' if it is 'sharp', *viz.* it is the most detailed one amongst all other images. This observation led to various *heuristic contrast-detection algorithms*; *cf.* [4,5], and the following functions – amongst others – were proposed as the sharpness index:

- the sum (or maximum) of absolute values (or squares) of differences of adjacent pixels,
- the amplitude of pixel brightness (*viz.* the span of a histogram brightness), or
- the image (or histogram) entropy.

The other (but equivalent) observation that 'the sharper image, the larger amplitudes of higher frequencies in the image Fourier transform' was used and formally examined in [6,7]. Also in [7], the use of the image variance as the focus function was mentioned. Eventually, an application of the continuous wavelet transform was proposed in [8].

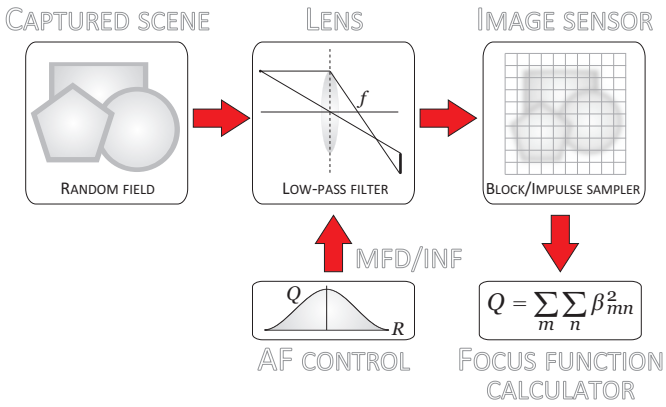

Fig. 1. The block diagram of the considered AF system

The algorithm proposed in the paper exploits the equivalencies between the image variance and the image orthogonal series expansion and is based on the observation that the focus function can quickly be evaluated using the orthogonal series transform and thus can readily be available in all digital camera devices equipped with the embedded transform coders like the JPEG, [9], JPEG 2000, [10], or JPEG XR, [11].

II. Problem statement and the AF algorithm

The formal background of the approach is thoroughly motivated in the ongoing paper, [12], where the theoretical foundations and properties of the algorithm are examined. Here we shortly recollect the algorithm assumptions, *cf.* [6,4,7,5,8] and see Fig. 1:

1) The **scene** is a *two dimensional homogenous second-order stationary process* (thus an *ergodic (in the wide*

sense) random field) with unknown distribution and correlation functions; *cf.* [13].

2) The **lens assembly** is modeled with the help of the *first-order optics laws, cf.* [14], that is, the lens acts as a *simple centered moving average* filter with the order proportional to the distance of the sensor plane from the image plane and to the size of the lens aperture.
3) A square (*i.e.* two dimensional) **image sensor** is modeled either by:
   a) the *block sampler*, or
   b) by the *impulse sampler* preceded by a low pass (*AA – anti-aliasing*) filter.

Both sensor models in Assumption 3 are inspired by the devices widely used in digital photography. The *block sampler* approximates the Foveon-type sensors, in which a single pixel consists of three pairs of stacked color filters and corresponding sensors, see *e.g.* [15]. The *impulse sampler,* combined with the *low-pass filter*, corresponds to the widespread *Bayer's Color Filter Array* combined with the optical *AA-filter*, where the single image pixel is reconstructed from the sensor pixels via the special interpolation procedure called *demosaicing*; *cf.* [16,17,18]. In both cases the lens-produced image is assumed to be *orthogonally projected* onto the respective function subspace. The block sampler projects the image onto the space of piecewise constant functions, while the impulse sampler preceded by the low-pass filter projects the image onto the space of band-limited functions.

Remark 1: The continuous image is projected onto the respective space and gets the approximate discrete representation in a given basis. Clearly, due to natural physical limitations, the sensor captures only a finite (square in our case) part of the scene.

The natural bases on squares are constructed as the tensor products of the one dimensional bases on the intervals, and in case of the block sampler they are piecewise-constant *Haar* and *Walsh-Hadamard* bases. In case of the band-limited functions, the basis constituted by the sinc functions is usually replaced by the Fourier basis, sine, or cosine bases. Note however that, regardless of the sensor type (and the resulting basis), we can treat this representation as the matrix of pixels values and decompose the matrix using any discrete series orthogonal on two-dimensional interval.

A. Algorithm

The crucial for the algorithm is the observation that the image variance can serve as the focus function and that this variance can be approximated by orthogonal expansion coefficients evaluated from the image acquired by the sensor; see Lemma 1. The variance is the largest when the image is 'in-focus' and the AF algorithm is just an algorithm searching for the maximum of such function. Various (bi-)orthogonal expansions can be used; *cf.* [19,20,21,22,23,24,25,26]:

- trigonometric, based on *Fourier*, cosine, sine, *Hartley*, *Walsh-Hadamard* series,
- wavelet, *e.g.*, orthogonal *Haar*, *Daubechies*, biorthogonal *LeGall (5/3)* and *Cohen-Daubechies-Vial (9/7)* series, or
- polynomial, *e.g.*, *Chebyshev, Legendre* – or in general – any *Jacobi* family of discrete orthogonal polynomials series.

These expansions can be fast computed by the existing transform coders (implementing transform coding compression schemes; see *e.g.* [9,27,28,29,30,31,32,10,33,34]). The following discrete orthogonal series transforms used in the available transform coders are thus considered and compared in the paper:

- *Cosine* transform (performed in **JPEG** coder),
- *Haar wavelet* transform (employed in **JPEG 2K (Part II)** coder, and
- *Walsh-Hadamard* transform (used in **JPEG XR** standard).

The **JPEG** engine is applied in its standardized transformation of the 8x8 image subblocks, and moreover, used to implement a hierarchical (progressive) DCT (H-DCT) transform, where the DC components of the 8x8 subblocks are combined in the 8x8 macroblocks and treated as inputs in the another 8x8 DCT transform step. The **JPEG 2K** transform is performed at various levels – from the maximum level in the 'full transform', to the single-level one (note that the latter essentially amounts to the classic *contrast-detect* algorithm). The *Walsh-Hadamard* used in **JPEG XR** standard is mimicked by the *Haar* transform performed at 16x16 subblocks (being 4x4 macroblocks built upon 4x4 subblocks).

B. Algorithm properties

The following lemma describes the formal justification of the proposed algorithm.

Lemma 1: Under Assumptions **1-3**, the variance of the sensor image is *unimodal* function with respect to the order of the lens filter model and attains its maximum value for the in-focus image.

Proof: Using the *first-order optics laws*, one can easily ascertain that the image in the image plane (*i.e.* the image before sampling by sensor) is the output of the linear *simple centered moving average* filter (*viz.* the lens) driven by the scene image. Then, the image, described by the convolution of the scene image with the filter impulse response, is sampled by the sensor. The order R of the moving average filter (counted in pixels of the image sensor) can be determined from the illustration in Fig. **??**.

Assume that the lens aperture is circular. Then the image of the point light source at the scene is the uniformly filled circle of the radius proportional to the lens aperture diameter D and the distance $|s-v|$ between the image plane and the sensor plane; *cf. e.g.* [5]:

$$R \sim D \cdot |s-v| \tag{1}$$

Clearly, the image is 'in-focus' when $s=v$. Let now

$$N(R) = 1 + 4\sum_{r=0}^{R} \left\lfloor \sqrt{R^2 - r^2} \right\rfloor \tag{2}$$

be a number of square pixels with centers inside the boundary of a circle of radius R. Let also

$$n(R) = N(R) - N(R-1)$$

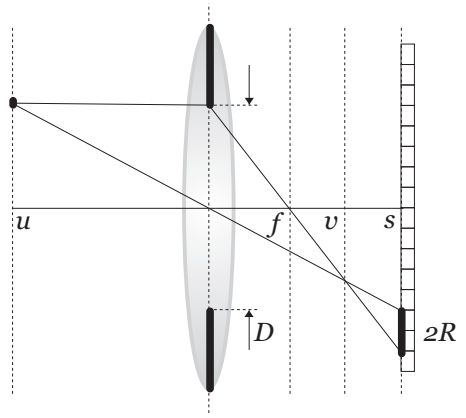

Fig. 2. Geometric construction of the image based on first-order optics laws. The lens is governed by the well-known relation $1/u+1/v=1/f$, where f is the lens focal length, u, v, s are, respectively, the distances of the scene, of the in-focus image, and of the image sensor from the lens. D is the aperture diameter

be a number of pixels placed on the circumference of this radius.

By virtue of Assumptions 1-3 (*i.e.* due to stationarity of the scene and time-invariance of the lens and sensor) we can consider a single pixel of the whole image $Y(x,y)$, $x=1,\ldots,I_X$, and $y=1,\ldots,I_Y$. For simplicity, we will denote this pixel as $Y_{00}=Y(0,0)$ and the whole image as Y. For the radius R, the sensor pixel, Y_{00}, is a (weighted by $N(R)$) sum of $N(R)$ image pixels $X_{r\varphi}$ grouped in $R+1$ circumferences of radii $r=0,\ldots,R$ (each circumference consists of $n(r)$ points)

$$Y_{00}=\frac{1}{N(R)}\sum_{r=0}^{R}\sum_{\varphi=n(0)}^{n(r)}X_{r\varphi}$$

The variance of the single pixel is thus (assuming for simplicity that the input process is centered, *i.e.* that $EX_{i\varphi}=0$):

$$\begin{aligned}\operatorname{var}Y_{00} &= \tfrac{1}{N^2(R)}\operatorname{var}\left(\sum_{r=0}^{R}\sum_{\varphi=n(0)}^{n(r)}X_{r\varphi}\right)\\ &= \tfrac{1}{N^2(R)}\sum_{i=0}^{R}\sum_{\varphi=n(0)}^{n(i)}\sum_{j=0}^{R}\sum_{\varphi=n(0)}^{n(j)}E\left(x_{i\varphi}x_{j\psi}\right)\\ &= \tfrac{1}{N^2(R)}\sum_{i=0}^{R}\sum_{j=0}^{R}\sum_{\varphi=n(0)}^{n(i)}\sum_{\varphi=n(0)}^{n(j)}\rho(i,j,\varphi,\psi)\end{aligned} \quad (3)$$

The pivotal for the algorithm is the observation that the variance of the whole image is (in spite of the correlation structure) a simple sum of variances of each pixels. This allows us to consider the (simpler) problem of unimodality of the variance of the single pixel image.

Remark 2: The fact that the variance of the image is simple a sum of variances of all pixels can easily be shown using the orthogonal representation of the image. Let $\{\varphi_{mn}(x,y)\}$, $m,n=0,1,\ldots$ be the two-dimensional discrete orthogonal basis on the square and let $\beta_{mn}=\langle Y,\varphi_{mn}\rangle$ be the coefficients in this expansion. Then, clearly

$$Y(x,y)=\beta_{00}\varphi_{00}(x,y)+\sum_{m=1}^{I_X}\sum_{n=1}^{I_Y}\beta_{mn}\varphi_{mn}(x,y). \quad (4)$$

By definition, the image squared energy is the sum of squares of the orthogonal expansion coefficients, that is

$$Y^2(x,y)=\left[\beta_{00}\varphi_{00}(x,y)+\sum_{m=1}^{I_X}\sum_{n=1}^{I_Y}\beta_{mn}\varphi_{mn}(x,y)\right]^2$$

and, by *orthonormality* of the basis functions, *i.e.* by the fact that $\langle\varphi_{mn},\varphi_{m'n'}\rangle=\delta(m-m')\cdot\delta(n-n')$, it is equal to the sum of squares of the expansion coefficients

$$Y^2(x,y)=\beta_{00}^2+\sum_{m=1}^{I_X}\sum_{n=1}^{I_Y}\beta_{mn}^2 \quad (5)$$

If in the basis $\{\varphi_{mn}(x,y)\}$, the first function, $\varphi_{00}(x,y)$, is the constant one (which holds in *Fourier, Legendre, Haar and Walsh-Hadamard* bases), then the term β_{00}^2 can be interpreted as the squared mean value of the image. Hence, the remaining double sum is just the variance of the sensor image (and the approximation of the scene image convolved with the lens).

Combining together (3)-(5), one can ascertain that for any second-order *homogeneous* stationary process, that is, the process with the *autocovariance function* depending only on a distance between pixels:

$$\rho(i,j,\varphi,\psi)=\|X_{i,\varphi}-X_{j,\psi}\|_{l_p},1\le p\le\infty,$$

and vanishing with a growth of the distance, *i.e.*,

$$\rho(i,j,\varphi,\psi)\to 0 \text{ as } \|X_{i,\varphi}-X_{j,\psi}\|_{l_p}\to\infty,$$

the variance $\operatorname{var}Y_{00}$ (and *a fortiori* the variance of the image) is unimodal and has a maximum at $R=0$.

Remark 3: Noting that (i,φ) are coordinates of $x_{i,\varphi}$ in a *polar system* with the pole in $(0,0)$, the *Euclidian* distance between two points in the *Argand complex plane* can be computed as $\|x_{p,q}-x_{r,s}\|_2$where

$$\begin{aligned}p &= i\cos\kappa(i,\varphi) \text{ and } q=i\sin\kappa(i,\varphi)\\ r &= j\cos\kappa(j,\psi) \text{ and } s=j\sin\kappa(j,\psi).\end{aligned}$$

where $\kappa(i,\varphi)=2\pi\varphi/n(i)$, is the angle (in radians). Hence

$$\|x_{p,q}-x_{r,s}\|_2=\sqrt{(p-r)^2+(q-s)^2}.$$

Example 1: In the simplest case of the *white noise* input process we have

$$\rho(i,j,\varphi,\psi)=\varrho\times\delta(i-j)\,\delta(\varphi-\psi)$$

and hence

$$\operatorname{var}y_{00}=\tfrac{1}{N^2(R)}\underbrace{\sum_{i=0}^{R}\sum_{\varphi=n(0)}^{n(i)}\varrho}_{N(R)\text{ pixels}}=\tfrac{\varrho}{N(R)}.$$

Example 2: In case of a *flat wall* (which however is **not** a second order stationary process since ρ does not vanish) we have in turn that $\rho(i,j,\varphi,\psi)\equiv\varrho$ and thus

$$\operatorname{var}Y_{00}=\text{const}\Longrightarrow\operatorname{var}Y=\text{const}$$

and the algorithm does not converge. Note however that it such a case the image is *always in focus!*

Remark 4: The image captured by the sensor is only a fragment of the scene, and (which is – in fact – well known in statistic literature) in case when the scene is a highly correlated process, the estimate of the variance can be very inaccurate and, in particular, does not resemble its unimodal origin.

C. AF criteria

In order to systematically describe the properties of the proposed focus function, we discuss the in the context of the focus function criteria presented in [4]:

1) **Unimodality**. This condition is fulfilled – as shown formally in the paper. The sufficient condition of unimodality requires only the autocorrelation function of the scene process to vanish with the distance. In practice, the unimodality can be lost in low-light situations because of the then-manifesting random character of the light due to the shot, thermal and quantization noises (resulting in the poor signal-to-noise ratio, *cf.* Fig. 5).
2) **Accuracy**. It depends on the resolution of the sensor. Since the same sensor is used in autofocusing and in capturing the image of interest, the accuracy is clearly the best attainable. In other words, since the captured image is the best approximation of the image yielded by the lens, then the focus function is the largest when the lens-produced image has the largest variance.
3) **Reproducibility**. A sharp top of the extremum is – in theory – the consequence of the fact that the relation between image variance and the order of the lens filter is reciprocal; see (3). In practice, the sharpness of the top can be attenuated by the influence of other (non-target) objects situated on the scene image.
4) **Range**. The variance of the image does not vanish for any finite $R \sim |s - v|$ and this theoretically guarantees convergence from any initial position of the lens. In practice, however, the range is limited by the size of the sensor: if, *e.g.,* R is larger than the diameter of the lens aperture, then clearly, the unimodality is lost. This issue can be – to some extent – attenuated by reducing the lens aperture diameter; see the experimental results in Fig. 6; *cf.* also the results of the analysis performed in *Fourier* domain in [6].

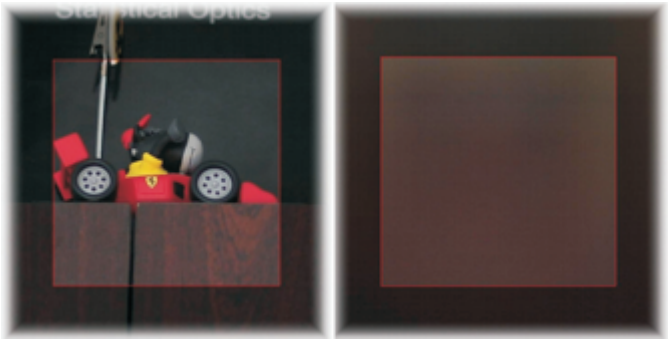

Fig. 3. Exemplary scene. Left – in focus, right – at minimum focus distance

5) **General applicability**. The generic class of processes is admitted. For instance, it covers all stable ARMA models, *Markov* fields, and *Cohen*'s PSM models; *cf.* [35]. The texture-rich images can be modeled, for example, by the uniformly distributed white noise process while the separated pointwise light sources process corresponds to the binomial or Poisson process. The *first-order optics law* seems to be sufficient to describe camera lenses – mainly because these lenses are carefully designed to be free of higher-order distortions; *cf.* [14]. Finally, both sensor models are close approximations of the two most popular sensor types.
6) **Insensitivity to other parameters**. Clearly, the variance is independent of the mean intensity of the image, and, for instance, any change of the mean brightness of the image (caused *e.g.* by the varying backlight) does not affect the focus index function. Note, however, that the scene objects situated in various distances from the lens can also affect unimodality of the focus function.
7) **Video signal compatibility**. The focus function is evaluated directly from the captured image. There is thus no disparity between the image and the data used for focusing. Such a disparity can rather occur when the focusing system is the autonomous one and exploits a separate optical/electric elements (like in *e.g.* single-lens reflex (SLR) cameras).
8) **Fast implementation**. The algorithms based on orthogonal expansions are usually fast – be it *e.g.* the *fast Fourier transform* and its real versions (*i.e.* DCT, or DST, amongst others), or the *fast Walsh* or *fast wavelet transform*. Furthermore, transform coders offers hardware implementations which are both speed and power consumption-optimized; *cf.* [10,36,37,38,39].

D. Experimental results

Several experiments have been performed to illustrate the accuracy and natural limitations of the proposed algorithm. In experiments, an assembly of the *Canon EOS* digital camera and two lenses with the focal lengths *85mm* and *100mm* were used. The camera was controlled by the application built upon the Canon *EDSDK* library. In the experiments, the algorithm was tested for:

- various lens aperture diameters,
- orthogonal series, and various
- AF region sizes.

The focus function was estimated as follows; *cf.* (4) and (5):

$$\bar{Q}_\beta = \log_2 \left[\sum_m \sum_n \bar{\beta}_{mn}^2 - \bar{\beta}_{00}^2 \right]$$

where $\{\bar{\beta}_{mn}\}$ are the empirical (acquired by the image sensor) image expansion coefficients calculated for a given discrete orthogonal basis (the $\log_2$ function was used only to limit the dynamic range of the focus function estimate. This is a one-to-one map and hence does not affect unimodality).

With an exception of the experiment presented in Fig. 4b, the AF region was always a 256x256 square. Note that **CD**

refers to the one-level Haar wavelet transform, **XR** to the four-level, and **MR-CD** to the maximum-level one.

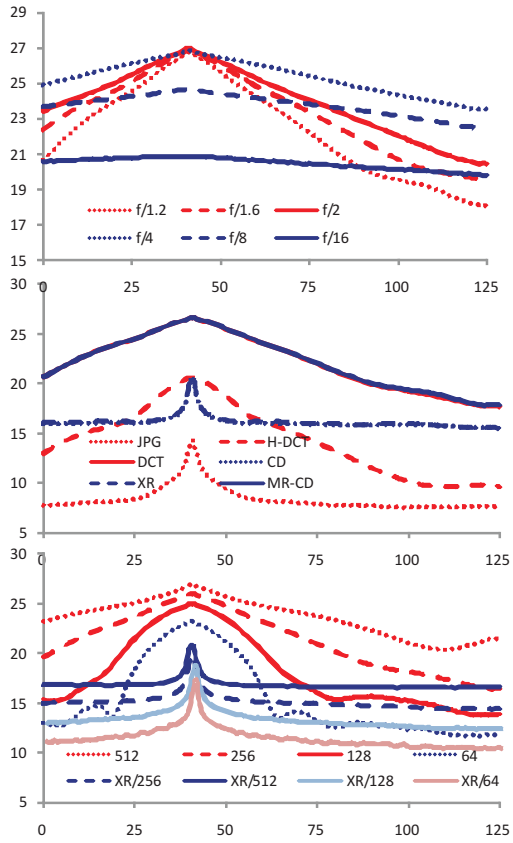

Fig. 4. Focus index for various: **(a)** aperture diameters (**JPEG2K**), **(b)** transform coders, **(c)** AF region sizes (**JPEG2K** vs **JPEG XR**)

The results confirm our formal findings; *cf.* the diagrams of the focus functions presented in Fig. 4. Below we shortly analyze selected factors affecting in practice the unimodality of the proposed focus function. To this end we consider the following two issues (*cf.* the previous section):

- the noise, and
- the boundary effect.

Presence of the noise is responsible for appearance of the small and random fluctuations in the focus function. The formal analysis of the noise sources is difficult since one needs to consider the combined impact of various noise types like *e.g.*: *photon shot, dark current, reset, thermal, quantization, and pattern* noises; *cf.* [40]. For instance, the shot noise is not the *i.i.d.* but depends on pixel values, *i.e.* its mean value (and variance) are both conditioned by the sensor pixel value.

Solution 1: The simplest countermeasure to this problem – based on the assumption that all the noise signals satisfy typical conditions of the *strong law of numbers* (*i.e.* they have, for a given pixel, the finite means and variances) – is an averaging routine in which the image, for a given position of the lens, is captured multiple times. This is the standard averaging technique and one can clearly expect that, with the growing number of repetitions, such routine cancels the random fluctuations more and more effectively (at the obvious cost of a larger computation overhead); see the results in Fig. 5.

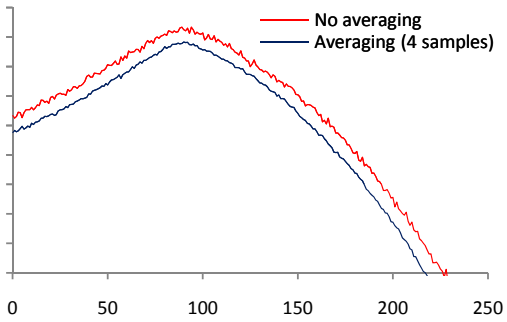

Fig. 5. Focus function evaluated from four- and from one-time sampled image

The boundary problem is responsible for the presence of nonrandom 'large blobs' on the focus function diagram, and is much more difficult to overcome (*cf.* the Remark 4 and the diagrams in Fig. 4c). Consider – as an illustrative example – a single pointwise source of light whose image is situated outside the image sensor. When the image is in-focus, then $R = 0$, and the captured image (or the focused region) is simply black (that is, it has zero energy and, *a fortiori*, zero variance). Otherwise, the image of the point becomes the circle of some radius R, *cf.* (1) and (2), and there exists some R' such that, for all $R \geq R'$, the circle intersects with the image sensor (its region) yielding a non-zero energy of the captured image and, in consequence, a non-zero variance and the focus function fails to have a maximum when the image is in focus (it can even have the local minimum there – see the Fig. 6!).

Solution 2: Recall that the radius R is proportional to the diameter D of the lens aperture. Thus, reducing the aperture we reduce the maximum radius and, in consequence, attenuate the influence of the 'boundary leakage' on the focus function values.[1] Note however, that the smaller aperture, the smaller signal-to-noise ratio.

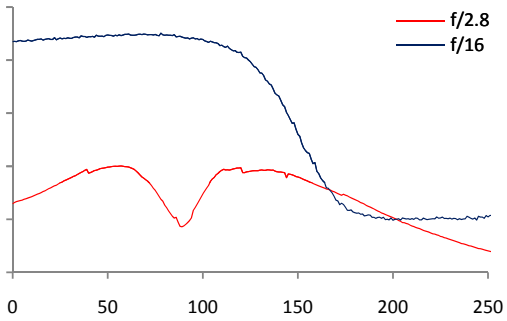

Fig. 6. Illustrations of the boundary effect manifesting in a form of 'light leakage' from the light sources outside the sensor. The non-unimodal focus function occurs here for $f/2.8$ aperture diameter. For the smaller aperture, $f/16$, the unimodality is recovered, however, at the price of the much larger noise

[1]The aperture diameter reduction is the technique already adopted in some newest digital SLR cameras like *Canon EOS 500D*.

E. Conclusions and final remarks

An application of the orthogonal series transforms, available in various transform coders, has been proposed and formally motivated for usage in AF algorithm. Efficiency of the approach in real-life tests has been confirmed. It remains, however, tempting (and easier to apply in practice) to exploit the whole transform coder rather than only its part performing the orthogonal transform, *i.e.* to employ the *entropy* of the image rather then its *variance* in the focus function machinery. This would allow, amongst others, using the *length* of the produced compressed stream as the focus function and employing the whole transform coder as the focus function calculator. The following observation may be helpful in derivation of the formal basis of this proposition.

Conjecture 2: Assume a *discrete zero-mean* random variable X; its *entropy* and *variance* are, respectively, given by the following well-known formulae:

$$H(X) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i \text{ and } \operatorname{var}(X) = \sum_{i=1}^{n} p_i x_i^2.$$

where $\{p_i = \Pr(X = x_i)\}_{i=1}^{n}$. Assume that p_i's are arranged in non-increasing order, *i.e.*, $p_i \leq p_j$ for $i > j$. Then, for any distribution of X such that $x_i^2 \geq -c \log_2 p_i$, some $c > 0$, the variance grows along with the growing entropy, which is, in turn, estimated by the size of the output stream produced by the transform coder.

Example 3: In a special case of the *uniformly distributed* X we simply have

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \cdot \log_2 \frac{1}{n} = \log_2 n$$

and

$$\operatorname{var}(X) = \frac{1}{6}(2n+1)(n+1)$$

that is, the *variance of* X grows with n. That the *entropy* grows is merely its natural property.

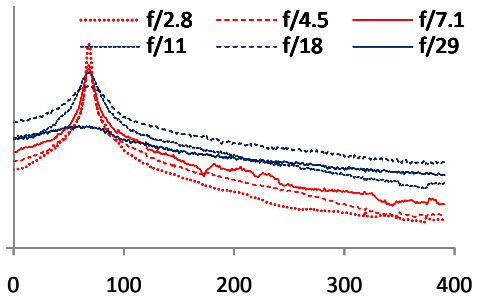

Fig. 7. The JPG filesize as the focus funtion against various aperture diameters, $D = f/2.8, \ldots, f/29$, $f = 100$. All maxima (*viz.* the largest sizes of JPG files) correspond to the in-focus images

To illustrate the conjecture we used the standard (lossy) JPEG coder and to measure the size of the coded (output) stream we simply used the size of the *.jpg* file. The results are presented in Fig. 7 and support the conjecture. We would like to emphasize that this algorithm can directly be used in almost all off-the-shelf cameras (*viz.* without *any* modification of the existing hardware and with only slight tweaking of the camera firmware).

References

[1] "Special issue on color image processing," *IEEE Signal Processing Magazine*, vol. 22, no. 1, 2005.

[2] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[3] K. S. Pradeep and A. N. Rajagopalan, "Improving shape from focus using defocus cue," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1920–1925, 2007.

[4] F. C. A. Groen, I. T. Young, and G. Ligthart, "A comparison of different focus functions for use in autofocus algorithms," *Cytometry*, vol. 6, no. 2, pp. 81–91, 1985. [Online]. Available: dx.doi.org/10.1002/cyto.990060202

[5] M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 864–870, 1998.

[6] A. Erteza, "Depth of convergence of a sharpness index autofocus system," *Applied Optics*, vol. 16, no. 8, pp. 2273–2278, 1977.

[7] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, 1987.

[8] J. Widjaja and S. Jutamulia, "Wavelet transform-based autofocus camera systems," *IEEE Proceedings*, 1998.

[9] W. B. Pennebaker and J. L. Mitchel, *JPEG: Still Image Data Compression Standard*. New York: Van Nostrand Reinhold, 1992.

[10] D. Taubman and M. Marcellin, *JPEG2000. Image Compression Fundamentals, Standards and Practice*, ser. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 2002, vol. 642.

[11] F. Dufaux, G. Sullivan, and T. Ebrahimi, "The JPEG XR image coding standard," *IEEE Signal Processing Magazine*, vol. 26, no. 6, pp. 195–199, 204, 2009.

[12] P. Śliwiński, "Autofocusing with orthogonal series," 2009, submitted for review.

[13] J. W. Goodman, *Statistical optics*. New York: Willey-Interscience, 2000.

[14] S. J. Ray, *Applied Photographic Optics, 3rd Ed.* Oxford: Focal Press, 2002.

[15] L. D. Paulson, "Will new chip revolutionize digital photography?" *IEEE Computer*, vol. 35, no. 5, pp. 25–26, 2002.

[16] R. Ramanath, W. E. Snyder, Y. Yoo, and M. S. Drew, "Color image processing pipeline. a general survey of digital still camera processing," *IEEE Signal Processing Magazine*, vol. 22, no. 1, pp. 34–43, 2005.

[17] D. D. Muresan and T. W. Parks, "Demosaicing using optimal recovery," *IEEE Transactions on Information Theory*, vol. 14, no. 2, pp. 267–278, 2005.

[18] X. Li, "Demosaicing by successive approximation," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 370–379, 2005.

[19] A. Haar, "Zur Theorie der Orthogonalen Funktionen-Systeme," *Annals of Mathematics*, vol. 69, 1910.

[20] G. Sansone, *Orthogonal Functions*. New York: Interscience, 1959.

[21] T. Rivlin, *Chebyshev Polynomials*. New York: Wiley, 1974.

[22] G. Szego, *Orthogonal Polynomials*, 3rd ed. Providence, R.I.: American Mathematical Society, 1974.

[23] M. Vetterli and D. Le Gall, "Perfect reconstruction FIR filter banks: some properties and factorizations," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 1057–1071, 1989.

[24] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM Edition, 1992.

[25] G. G. Walter, *Wavelets and other orthogonal systems with applications*. Boca Raton: CRC Press, 2001.

[26] M. Unser and T. Blu, "Mathematical properties of the JPEG2000 wavelet filters," *IEEE Transactions on Image Processing*, vol. 12, no. 9, pp. 1080–1090, September 2003.

[27] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, 1992.

[28] R. A. DeVore, B. Jawerth, and B. Lucier, “Image compression through wavelet transform coding,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 719–746, 1992.
[29] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, “Data compression and harmonic analysis,” *IEEE Transactions on Infromation Theory*, vol. 44, no. 6, pp. 2435–2476, 1998.
[30] “Special issue on JPEG2000 standard,” *IEEE Signal Processing Magazine*, vol. 18, no. 5, 2001.
[31] “Special issue on JPEG 2000 standard,” *Signal Processing: Image Communication*, vol. 17, 2002.
[32] A. Cohen, I. Daubechies, O. G. Guleryuz, and M. T. Orchard, “On the importance of combining wavelet-based nonlinear approximation with coding strategies,” *IEEE Transactions on Information Theory*, vol. 48, no. 7, pp. 1895–1921, 2002.
[33] “Special section on JPEG 2000 digital imaging,” *IEEE Transactions on Consumer Electronics*, vol. 49, pp. 771–888, 2003.
[34] “Special section on the H.264/AVC video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 557–725, 2003.
[35] A. Cohen and J.-P. D’Ales, “Nonlinear approximation of random functions,” *SIAM Journal of Applied Mathematics*, vol. 57, no. 2, pp. 518–540, 1997.
[36] T. Acharya and P.-S. Tsai, *JPEG2000 Standard for Image Compression: Concepts, Algorithms and VLSI Architectures*. Wiley-Interscience, 2005.
[37] D. T. Lee, “JPEG 2000: Retrospective and new developments,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 32–41, 2005.
[38] H.-C. Fang, Y.-W. Chang, C.-C. Cheng, and L.-G. Chen, “Memory efficient JPEG 2000 architecture with stripe pipeline scheduling,” *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4807–4816, 2006.
[39] G. Savaton, E. Casseau, and E. Martin, “Design of a flexible 2-D discrete wavelet transform IP core for JPEG2000 image coding in embedded imaging systems,” *Signal Processing*, vol. 86, pp. 1375–1399, 2006.
[40] B. Fowler, M. D. Godfrey, and S. Mims, “Reset noise reduction in capacitive sensors,” *IEEE Transaction on Circuits and Systems-I: Regular Papers*, vol. 53, no. 8, pp. 1658–1669, 2006.