Interest Points as a Focus Measure in Multi-Spectral Imaging

Martin ZUKAL, Jiri MEKYSKA, Petr CIKA, Zdenek SMEKAL

Dept. of Telecommunications, Brno University of Technology, Technická 12, 616 00 Brno, Czech Republic

martin.zukal@phd.feec.vutbr.cz, j.mekyska@phd.feec.vutbr.cz, cika@feec.vutbr.cz, smekal@feec.vutbr.cz

Abstract. *A novel multi-spectral focus measure that is based on algorithms for interest point detection, particularly on the FAST (Features from Accelerated Segment Test), Fast Hessian and Harris-Laplace detector, is described in this paper. The proposed measure methods are compared with commonly used focus measure techniques like energy of image gradient, sum-modified Laplacian, Tenenbaum's algorithm or spatial frequency when testing their reliability and performance. The measures have been tested on a newly created database containing 420 images acquired in visible, near-infrared and thermal spectrum (7 objects in each spectrum). Algorithms based on the interest point detectors proved to be good focus measures satisfying all the requirements described in the paper, especially in thermal spectrum. It is shown that these algorithms outperformed all commonly used methods in thermal spectrum and therefore can serve as a new and more accurate focus measure.*

Keywords

Focus measure, interest points detector, multi-spectral imaging, gradient, Fast Hessian, Harris-Laplace detector.

1. Introduction

An image fusion system is a system that combines information from two (or more) images into a single, more informative, image. The input images can be acquired in the same spectrum or in different spectra. An example of image fusion system that operates in visual spectrum can be multi-focus image fusion [1] which combines images acquired with different focus into a single image containing blocks that are all well focused. Common cameras have only limited depth of field which means that only objects within the depth of field are focused. Images produced by the multi-focus image fusion system have no limitations in terms of depth of field and thus convey more information than regular images.

There exist also systems that operate in more spectra. Human eye is able to perceive electromagnetic (EM) radi-ation of wavelengths between 300 nm and 700 nm. The response in this range is not flat and is described by the photopic curve which has a peak at 550 nm (light green color) where the human visual system is most sensitive. This portion of EM spectrum is referred to as the visual spectrum (VIS). The infrared spectrum which includes EM radiation of wavelength from 1 mm to 700 nm lies left from the visible spectrum. This part of the spectrum can be further divided into: near infrared (NIR) spectrum ($1.4\,\mu$m $- 700$ nm), medium infrared (MIR) spectrum ($3\,\mu$m $- 1.4\,\mu$m) and far infrared (FIR) spectrum (1 mm $- 3\,\mu$m). The thermal infrared (TIR) spectrum is a sub-band of FIR spectrum ranging from $3\,\mu$m to $14\,\mu$m. This radiation is perceived by humans in the form of heat.

It has already been proven that processing of images in NIR and TH spectrum is beneficial. These images contain unique information that is not available in visual spectrum and it is useful to combine this information [2]. Espinosa-Duró et al. showed that images in NIR spectrum have higher entropy than VIS and TH images and moreover according to the results of mutual information it can be said that the information in all spectra is not overlapping (each spectrum has a unique part of information) [2]. This is useful especially in multi-spectral image fusion system that combines images from visual, thermal and/or near infra red spectrum [3]. It has been shown that face recognition based on this system provides higher recognition rate than a system based on single-spectral images [4].

A focus measure plays an important role in the fusion system because it directly influences the performance of the system [1]. While autofocusing has been deeply studied in the visual spectrum, notably less work has been done on determining the optimal focus position in thermal images. There are, however, two studies that evaluate usefulness of different focus measures in thermal images [5] and its application in the image fusion systems [6]. Inspired by these studies, we propose a new focus measure which is primarily targeted on images in NIR and TH spectrum. This focus measure is based on interest points detected in an image and it is expected to have some good properties like invariance to scale, rotation, lightening conditions [7] etc. The use of an interest point detector as a focus measure for multi-spectral images is – to our best knowledge – novel and has never been

published before. In this paper, we compare this measure with the measures presented in the aforementioned study on a database that we created especially for this purpose.

The paper is organized as follows: focus measures are described in detail in Section 2, our proposed focus measure methods are explained in Section 3, Section 4 describes the experiments and Section 5 is devoted to the experimental results. The conclusion is given in Section 6.

2. Focus Measures

A focus measure, as defined in literature [8], should attain a maximum for the best focused image and decrease with increasing blur of the image. One can use the following experiment to evaluate the focus measure [9]. An object is placed in front of the camera which acquires one image at each lens position. The focus measure is computed for each image from the set described above. If the focus measure satisfies the condition described at the beginning of this paragraph, it will result in a curve (referred to as the sharpness curve in what follows) which will have a peak at the position of the best focused image.

A typical focus measure should meet the following requirements [5]:

1. It should not depend on the content of the image.

2. The focus measure should be unimodal, that is, it should have one and only one maximum value.

3. It should have large variation in value with respect to the degree of blurring.

4. The focus measure should have minimal computation complexity.

5. It should be robust to noise as much as possible.

There exist many focus measures such as Energy of image gradient, Sum-modified Laplacian, Tenengrad, Energy of Laplacian of the image and many others [5]. The ones that we used in our experiment will be described in the following section.

2.1 Variance

One of the simplest methods is to calculate variance (VR) in brightness using a small neighborhood at each pixel location [10]. The variance steeply increases for patches that are in focus. This simple measurement can be expressed by equation:

$$c_{VR} = \frac{1}{M \cdot N} \sum_{x=1}^{M} \sum_{y=1}^{N} [I(x,y) - \mu]^2, \qquad (1)$$

$$\mu = \frac{1}{M \cdot N} \sum_{x=1}^{M} \sum_{y=1}^{N} I(x,y). \qquad (2)$$

2.2 Energy of the Image Gradient

The energy of an image gradient (EOG) is based on the vertical and horizontal gradients of the image and is obtained according to equations:

$$c_{EOG} = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} (f_x^2 + f_y^2), \qquad (3)$$

$$f_x^2 = I(x+1,y) - I(x,y), \qquad (4)$$

$$f_y^2 = I(x,y+1) - I(x,y). \qquad (5)$$

2.3 Tenenbaum's Algorithm (Tenengrad)

Tenengrad (TEN) is based on a calculation of image gradient. This calculation is derived from Sobel operator:

$$c_{TEN} = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} [\nabla S(x,y)]^2 \text{ for } \nabla S(x,y) > T \qquad (6)$$

where T is a selected threshold and $\nabla S(x,y)$ is Sobel gradient expressed as:

$$\nabla S(x,y) = \sqrt{\nabla S_x^2(x,y) + \nabla S_y^2(x,y)}, \qquad (7)$$

$$\begin{aligned} \nabla S_x^2(x,y) = & -[I(x-1,y-1) + 2I(x-1,y) \\ & + I(x-1,y+1)] + [I(x+1,y-1) \\ & + 2I(x+1,y) + I(x+1,y+1)], \qquad (8) \end{aligned}$$

$$\begin{aligned} \nabla S_y^2(x,y) = & -[I(x-1,y-1) + 2I(x,y-1) \\ & + I(x+1,y-1)] - [I(x-1,y+1) \\ & + 2I(x,y+1) + I(x+1,y+1)]. \qquad (9) \end{aligned}$$

This measure is similar to EOG, however TEN uses Sobel's central differences instead of first-order left difference when approximating the first derivative.

2.4 Energy of Laplace Operator

Next focus measure technique is an energy of Laplace operator (EOL) given by formula:

$$c_{EOL} = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} [f_{xx} + f_{yy}]^2, \qquad (10)$$

$$\begin{aligned} f_{xx} + f_{yy} = & -I(x-1,y-1) - 4I(x-1,y) \\ & -I(x-1,y+1) - 4I(x,y-1) \\ & +10I(x,y) - 4I(x,y+1) \\ & -I(x+1,y-1) - 4I(x+1,y) \\ & -I(x+1,y+1). \qquad (11) \end{aligned}$$

2.5 Sum-modified Laplacian

Sum-modified Laplacian (SML) is derived from energy of Laplacian of the image. Nayar et al. [11] noted that in the case of the Laplacian, the second derivatives in the x and y

directions can have opposite signs and tend to cancel each other out. Therefore they proposed the sum modified Laplacian, which can be obtained according to formula:

$$c_{\text{SML}} = \sum_{i=x-W}^{x+W} \sum_{j=y-W}^{y+W} \nabla_{\text{ML}}^2 f(i,j) \text{ for } \nabla_{\text{ML}}^2 f(i,j) \geq T, \quad (12)$$

$$\nabla_{\text{ML}}^2 f(x,y) = |2I(x,y) - I(x-1,y) - I(x+1,y)| + $$
$$+ |2I(x,y) - I(x,y-1) - I(x,y+1)|. \quad (13)$$

In order to accommodate for possible variations in the size of texture elements, Nayar et al. used a variable spacing (step) between the pixels to compute ML. The parameter W determines the window size used to compute the focusing measure.

2.6 Spatial Frequency

Spatial frequency (SF) is a modified version of image gradient and is given by: [12]

$$c_{\text{SF}} = \sqrt{f_{\text{r}}^2 + f_{\text{c}}^2} \quad (14)$$

where f_{r} and f_{c} is a row frequency and a column frequency, respectively. These frequencies can be calculated according to:

$$f_{\text{r}} = \sqrt{\frac{1}{M \cdot N} \sum_{x=1}^{M} \sum_{y=2}^{N} [I(x,y) - I(x,y-1)]^2}, \quad (15)$$

$$f_{\text{c}} = \sqrt{\frac{1}{M \cdot N} \sum_{x=2}^{M} \sum_{y=1}^{N} [I(x,y) - I(x-1,y)]^2}. \quad (16)$$

3. Novel Focus Measure Based on the Number of Interest Points

We base our focus measure on the number of interest points detected in the image by an interest point detector. An interest point is a point in an image where certain property changes significantly. Probably the most frequently considered property is intensity [13]. It can be observed that the number of detected interest points decreases when the image is blurred. It does not matter whether the algorithm detects corners, blobs or regions in general. All these features have to be well defined in the image space in order to be detected. Blurring the image significantly lowers the possibility of detecting the interest points. Therefore, we formulate the focus measure as follows:

$$c_{\text{IP}} = \frac{n - n_{\min}}{n_{\max} - n_{\min}} \quad (17)$$

where n is the number of points detected in the currently processed image, $n_{\max}$ denotes the maximum number of interest points detected in the set of images of one object and $n_{\min}$

denotes the minimum number of interest points detected in the set of images of one object. Typical values of the number of detected points are discussed in Section 5. The used implementation of the algorithms for interest point detection allows us to limit (with a parameter) the number of the detected points to a subset of best points or to use all detected points. We used the latter option since the number of points is strongly dependent on the scene and is not possible to be estimated beforehand.

There has been a great number of interest point detectors proposed in literature over the past years. We selected three state-of-the-art interest point detectors and included them in our experiment. Specifically, we used Fast Hessian detector, the FAST (*Features from Accelerated Segment Test*) detector and the Harris-Laplace detector. The first two are implementations that were designed to be quick to compute while the Harris-Laplace is not optimized for speed. The Fast Hessian and Harris-Laplace detector both detect points that are scale invariant whereas the FAST detector detects points in one scale only.

3.1 FAST

The FAST algorithm [14] considers a Bresenham circle (of radius 3) of pixels around the candidate pixel. This circle consists of sixteen pixels (see Fig. 1). The candidate pixel p is marked in yellow color and the pixels in the circle are displayed as blue ones and numbered as depicted in the figure. The intensity of the candidate point is denoted as I_p. The detector classifies p as a corner if there exist at least 12 contiguous pixels in the circle which are all brighter than $I_p + t$ or all darker than $I_p - t$ where t is a defined threshold which is a parameter of the algorithm.
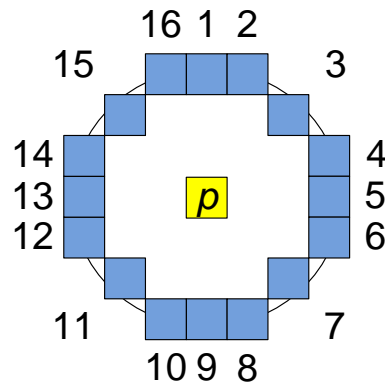

Fig. 1. Candidate corner and surrounding pixels.

The following conditions are result of training of a decision tree classifier to detect corners. The first condition considers pixels p_1 and p_9. If pixels

$$I_p - t \ < p_1 < \ I_p + t \quad (18)$$

or

$$I_p - t \ < p_9 < \ I_p + t \quad (19)$$

the pixel p is not a corner.

Otherwise the pixels p_5 and p_1 are tested. If p should be selected as a corner, then the equations

$$p_x \quad < \quad I_p - t \tag{20}$$

and

$$p_x \quad > \quad I_p + t \tag{21}$$

where p_x is selected pixel, are true for at least three of tested pixels.

More detailed description of this method can be found in [14].

3.2 The Fast Hessian Detector

The Fast Hessian (FH) detector [15] is based on the Hessian matrix. The Hessian matrix at a point $\mathbf{x} = (x, y)$ in an image I can be described in the following way

$$\mathbf{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \tag{22}$$

where $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative in the x dimension with the image I in the point $\mathbf{x}$, $L_{yy}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative in the y dimension with the image I in the point $\mathbf{x}$, and $L_{xy}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order derivative in the xy dimension with the image I in the point $\mathbf{x}$.

The second order derivatives are combined into a single measure using the determinant of the Hessian matrix. This measure gives weak response in points where the signal change is most significant. Conversely, the maximum is reached at ridges and blobs (roughly circular shapes that are darker or brighter than the surrounding). [13]

The calculation of the approximation of (22) exploits so called integral images [16] which speed up the computation a lot.

3.3 Harris-Laplace Detector

The Harris-Laplace (HL) detector [17] is a scale adapted modification of the Harris corner detector [18]. The Harris-Laplace utilizes the second order matrix which describes the structure of the local neighborhood of a point to detect interest points. The scale invariant second order matrix for point $\mathbf{x}$ is defined as

$$\begin{aligned} \mathbf{M}_L(\mathbf{x}, \sigma_I, \sigma_D) &= \sigma_D^2 g(\sigma_I) \mathbf{M}(\mathbf{x}, \sigma_D), \\ \mathbf{M}(\mathbf{x}, \sigma_D) &= \begin{bmatrix} L_x^2(\mathbf{x}, \sigma_D) & L_x L_y(\mathbf{x}, \sigma_D) \\ L_x L_y(\mathbf{x}, \sigma_D) & L_y^2(\mathbf{x}, \sigma_D) \end{bmatrix} \end{aligned} \tag{23}$$

The idea behind the equation is to average the local derivatives computed with Gaussian kernels of the size σ_D (differentiation scale) in the neighborhood of the point. One

effective way to compute the average is to convolve the image with a Gaussian window G of the size σ_I (integration scale).

Candidate points are localized as maxima of so called cornerness H in the first step of the algorithm. The cornerness is defined by the following equation

$$H \quad = \quad \det \mathbf{M}(\mathbf{x}, \sigma_I, \sigma_D) - k(\operatorname{trace} \mathbf{M}(\mathbf{x}, \sigma_I, \sigma_D))^2 \tag{24}$$

where $\det \mathbf{M}(\mathbf{x}, \sigma_I, \sigma_D) = L_x^2 L_y^2 - (L_x L_y)^2$, $\operatorname{trace} \mathbf{M}(\mathbf{x}, \sigma_I, \sigma_D) = L_x^2 + L_y^2$ and k is a constant.

The second step consists of computation of the Laplacian-of-Gaussian

$$|\operatorname{LoG}(\mathbf{x}, \sigma_n)| \quad = \quad \sigma_n^2 |L_{xx}(\mathbf{x}, \sigma_n) + L_{yy}(\mathbf{x}, \sigma_n)| \tag{25}$$

and selection of points that reach a maximum. These points are selected as interest points. For a detailed description of the algorithm with precise values of all variables refer to [17].

4. Experimental Setup

Focus measures are computed once the images are acquired. All algorithms for computing focus measures have been implemented in Java programming language. The reason for this is that we wanted to measure real-world performance of these algorithms. There exists a relatively new library (released in November 2011) that provides an effective representation of the images in computer memory and it also contains implementation of some basic as well as advanced image processing algorithms. The library is called BoofCV[1]. It was released under the Apache license and it is free for academic use. It has been decided to use the classes representing the image from this library and to program all missing algorithms (the focus measures not based on interest point detectors). The implementation of interest point detectors from the BoofCV library was used in our experiment. The focus measures were computed one hundred times for each image. The time needed to compute the value was measured using a Java built-in method (System.nanoTime()). Basic statistics (mean and standard deviation) were computed for each algorithm in each set to obtain values that can be further compared. We are aware of the fact that this result is biased since the processor did not spend all its computing time working on this task but the one hundred times repetition gives a good estimate of the complexity of each algorithm. For general evaluation of the times we provide a time complexity analysis in Section 5. The experiment was conducted on a computer with Intel Core i5-2320 @ 3.00 GHz CPU with 16 GB RAM.

The focus measures are evaluated according to different criteria such as monotonicity magnitude of slope and smoothness. To compare the selected methods using a quantitative evaluation there was selected an easy measurement

---

[1] http://boofcv.org

based on a Q factor. This factor can be calculated according to formula:

$$Q = \frac{1}{n_{max} - n_{min} + 1}, \tag{26}$$

$$c_s[n] \geq 10^{-\frac{3}{20}} \text{ for } n_{min}, \ldots, n, \ldots, n_{max} \tag{27}$$

where $c_s[n]$ represents the discrete sharpness curve normalized to range [0;1]. This idea was adopted from a filter theory. We are counting number of sharpness curve samples that have values higher than 0.7079 ("attenuation" is lower than 3 dB) around the main peak. Generally a narrow peak in the sharpness curve (i. e. the curve generated by better method) has higher Q factor than the wider one.

4.1 Image Databases

We acquired three databases, each consisting of seven sets of images. One database contained images acquired in visual spectrum, one in near infrared spectrum and one in thermal spectrum.

We tried to make the scenes as variable as possible. The scenes contain usually one object but there are also scenes that contain more objects. Furthermore, the distance of the object from the camera as well as the area occupied by the object in the entire image varied in each set. The camera was put on a tripod in order to ensure that exactly the same object (scene) will be captured when shooting each set of images. As the thermal camera produces images with resolution 640 × 480 pixels, the images captured by other cameras were resized and cropped to 640 × 480 pixels in order to accommodate to the resolution of the thermal camera. The thermal camera captures a heat map of the scene which can be represented as an image with indexed color with 8 bits per pixel (grayscale image). Therefore the images acquired in visual and NIR spectrum were converted to grayscale before the experiment.

The visual spectrum database was acquired by Nikon D80 camera with 50 mm f/1.8D AF Nikkor lens. We moved manually the focusing ring in 5 mm steps. This resulted in 12 images that were captured in each set. We are aware of rather coarse step, but since we are more interested in images in NIR and TH spectrum we did not strive to refine the step. The scene was lit with Fomei photographic lights and softboxes. No additional lighting was required therefore we did not use flash light. An example consisting of five images selected from the "Mixer" set is depicted in Fig. 2.

The near infrared image database was acquired using Canon EOS 350D camera with Canon EF-S 18-55 mm f/3.5-5.6 lens. The camera was adjusted to capture the EM radiation in NIR spectrum by removing the infrared filter before the CMOS image sensor. We mounted a B+W 093 infrared filter on the lens. This filter blocks all visible light and allows only infrared light above 850 nm to pass. The cut wavelength was measured using VARIAN CARRY 50 UV-VI spectrophotometer. We moved manually the focus-

ing ring in 1 mm steps in this experiment which resulted in 21 images in each set. We used the built-in flash to light the indoor scenes. The outdoor scenes were shot without additional lighting. An example consisting of images selected from the "Engine" set is depicted in Fig. 3.

The thermal image database was acquired by a thermal camera FLIR SC660. This camera is equipped with an uncooled detector and has a spectral sensitivity range from 7.5 to 13 μm. It provides image resolution: 640 × 480 px; optical field/min. focus distance: 45° × 34° / 0.2 m; thermal sensitivity: 45 mK at 30 °C and spatial resolution: 0.65 mrad. More information about this camera can be found in datasheet [19]. An example consisting of five images selected from the "Head" set is depicted in Fig. 4.

The described database is freely available provided that any publications based on this database will cite this paper. The database can be downloaded at http://splab.cz/en/download/databaze/multispec.

5. Experimental Results and Discussion

The results of all measurements can be seen in Tab. 2 and the resulting sharpness curves in Fig. 5 to Fig. 25. Each algorithm in particular spectrum was evaluated according to the Q factor, the shape of the sharpness curve $c_s[n]$ and the position of the maximum (P) of the sharpness curve. We also evaluated the best focused image subjectively and included this information in the table in the SUB column. Finally, mean and standard deviation of required time t to calculate the sharpness coefficient for each image is included in the table as well.

We have observed that the number of detected interest points depends on particular algorithm for detection of the interest points and the scene but we did not observe any correlation between the spectra. The values of n_{min} and n_{max} are gathered in Tab. 1 from which can be seen that the smallest number of interest points was detected by the Harris-Laplace detector, the greatest number of interest points detected the FAST algorithm and the Fast Hessian algorithm detected points whose number was between the FAST and Harris-Laplace algorithms.

	Fast Hessian		FAST		Harris-Laplace	
	n_{min}	n_{max}	n_{min}	n_{max}	n_{min}	n_{max}
min(n)	0	57	378	483	2	41
max(n)	161	379	836	950	80	176
mean(n)	29.90476	182.9048	623.381	782.3333	35.1	104.5

Tab. 1. Values of n_{min} and n_{max} for different Interest Point Detectors.

The discrimination power of the proposed measure directly depends on the range $n_{max} - n_{min}$. The average range was around 150 points for Fast Hessian and FAST algorithms and about 70 points in case of the Harris-Laplace detector. The lowest discrimination power has the Harris-

Fig. 2. Example of blurred and well focused images in "Mixer" (VIS) set. Sharp image is the central one.
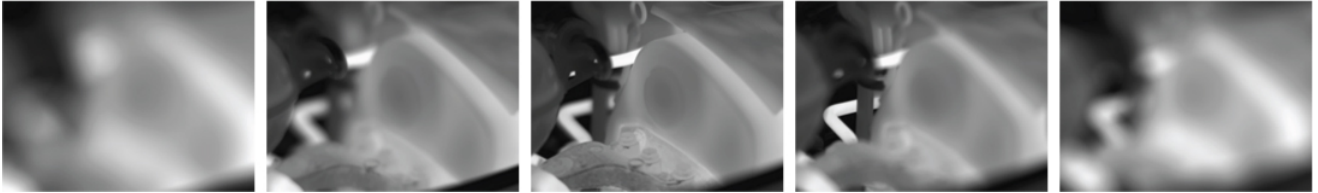

Fig. 3. Example of blurred and well focused images in "Engine" (TH) set. Sharp image is the central one.

Laplace detector which did not give very good results anyway (as discussed further). The discrimination power of the other two algorithms is satisfactory for our experiment.

5.1 Evaluation of Methods in the Visible Spectrum

According to the results it is obvious that the most narrow and high sharpness curve was calculated in all sets using the energy of Laplace operator. The second best results were provided by energy of image gradient then tenengrad, summodified Laplacian and spatial frequency. Methods based on the interest points detection did not provide good results. The derived sharpness curves are flat, there are a lot of changes in direction and it is hard to detect a global maximum. This contrasts with our previous results presented in [20], where we tested the focus measures on the same database. The method based on Fast Hessian gave comparable results to other focus measures in that study. After an analysis of the experiments, we can conclude that the difference is caused by a different approach that was used to convert the images to grayscale. We used directly the color images in our previous experiment and we extracted the luminance (Y) matrix according to the following formula

$$Y = 0.299R + 0.587G + 0.114B \qquad (28)$$

where R, G and B denote the red, green and blue image bands.

In the current experiment we used an external program (IrfanView) to perform this step. IrfanView uses the following formula to obtain the grayscale version (Y') of the image

$$Y' = 0.333R + 0.333G + 0.333B. \qquad (29)$$

This introduces rather significant difference in the pixel values. The interest point detectors are designed to be invariant to rotation, scale change and even affine transforms but they are not designed to be invariant to transformation that is introduced by multiplying the image bands with different weights.

Similar bad results were detected when evaluating sharpness curves based on the variance. In case of methods EOG, EOL, SML, SF and TEN, the selected position of maximum (P) was in all sets same as position selected subjectively. In case of FAST, there were 3 sets where the automatic selection was equal to the subjective one. When evaluating the methods according to Q factor, it can be said that EOL again provided the best results.

5.2 Evaluation of Methods in the Near-infrared Spectrum

Similarly to the visible spectrum, the best candidate for the image focus measurement is a method based on the energy of Laplace operator. This method provided the best results in all tested sets. Also energy of image gradient provided narrow and high sharpness curve in all sets. The third best candidate is SML or TEN depending on the particular set. In case of "Building", "Car" and "Corridor" sets the SML gave better results than TEN. The sharpness curve based on TEN has usually bigger distance between peak and valley, but the peak is wider than in case of SML. Narrow peak of the sharpness curve is considered better as the algorithm for automatic focus measurement quickly converges to the best position. Looking at the Q factor values, it can be said that TEN provided higher factor only in case of "Keyboard" set. This set is rather complicated for automatic focus because the object occupies large part of the image spreading from long to short distances. This means that there is always a part of the image that is well focused. The sharpness curve should attain maximum at the image which contains the largest well focused part. In fact, this set would be especially interesting for image fusion, where different well focused parts can be combined into one perfectly focused image. This is one of the topics for the future work. Methods FH, FAST, HL and VR provided very bad results. The results could be further improved by employing methods which would enhance the contrast of the image which is rather poor in case of NIR images. Such method was proposed in [21]. Probably worst-performing method was the
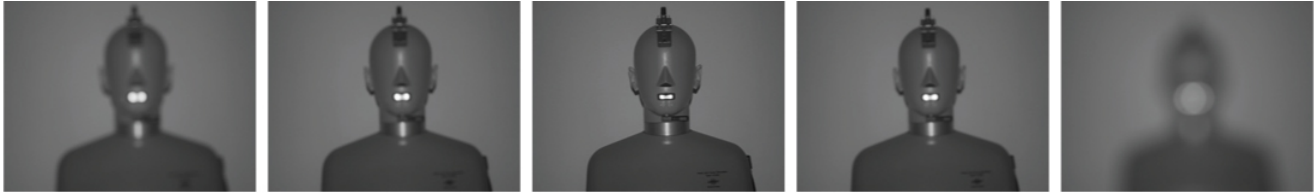
Fig. 4. Example of blurred and well focused images in "Head" (NIR) set. Sharp image is the central one.

one that uses the Harris-Laplace detector. There can be seen a noticeable valley close to the position of the well-focused image in case of HL method making this method inapplicable as a focus measure. Regarding the Q factor, it can be said that there were methods equal but not better than EOL.

5.3 Evaluation of Methods in the Thermal Spectrum

Contrary to the visible and near-infrared spectrum it can be said that the best results were obtained using method based on Fast Hessian detector. Sharpness curve calculated according to this detector is usually very narrow, high and with an evident global maximum. These results serve as a proof that the method based on the interest points detection can be used as a focus measure and moreover it can provide very good results in thermal spectrum. The second best method is TEN. The rest of the methods provided bad results at least in one set. In "Circuit Breaker", "Building", "Printer" and "Server" sets there can be also noticed significant changes in the course of function. Looking at the Q factor, it can be said that good results were provided by EOG and TEN in sets "Circuit Breaker", "Building" and "Server". In the rest of sets the best values were calculated in case of fast Hessian. Regarding the P values it can be said the TEN provided same results as SUB in all sets. Method FH was in two cases equal to SUB and in 5 cases different from SUB by ± 1 step.

5.4 Evaluation of the Computational Complexity

We will analyze the time complexity of the algorithms in terms of big O notation in this section. This analysis is a common tool that can be used to compare complexity of algorithms.

The algorithms operate over an image with dimension $m \times n$ pixels. As big O notation analyzes the worst case that can happen, we will consider n to be the larger value and we will consider square images of dimension $n \times n$ in what follows.

The regular algorithms (VR, EOG, TEN, EOL, SML and SF) are implemented as one-pass algorithms so they inspect each pixel only once. The time complexity of a one-pass algorithm is $O(n^2)$. The algorithms differ in the operations that needed to be performed on each pixel position. The number of operations for each pixel is different in each

algorithm but it is constant ($O(1)$) for each of the algorithms. The multiplicative constant can be ignored in the worst case analysis. Therefore we can say that the algorithm runs in

$$f_{\text{reg}}(n) \quad \in \quad O(1) \cdot O(n^2) \approx O(n^2) \qquad (30)$$

(quadratic) time.

The analysis of algorithms for interest point detection is more complicated. Firstly, the data structures to store the temporary data and the extracted interest points are initialized. This overhead takes constant time and it has to be taken into consideration for all algorithms based on interest point detection.

The simplest of the interest point detectors is the FAST algorithm. The core algorithm is one-pass algorithm but the number of operations for each pixel differs according to the nature of the image. Let us assume that the maximum (but constant) number of steps would be performed at each pixel position. After the candidate points are selected, a non-maximum suppression is performed which needs to inspect all points in the image, thus having complexity of $O(1) \cdot O(n^2)$ since a constant number of steps is performed at each pixel position. Putting this together yields:

$$
\begin{aligned}
f_{\text{FAST}}(n) \quad \in \quad & O(1) + O(1) \cdot O(n^2) \\
& + O(1) \cdot O(n^2) \\
\approx \quad & O(n^2).
\end{aligned}
\qquad (31)
$$

The quadratic function dominates the constant one therefore the complexity of the Fast Hessian detector is $O(n^2)$.

The Fast Hessian detector operates with integral images that are computed in the first step of the algorithm. This can be done with $O(1) \cdot O(n^2)$ complexity. The image is processed in multiple octaves (4 in our case). The algorithm exploits the integral image to compute the Hessian matrix in each octave. The analysis of the algorithm that is used to compute the Hessian matrix is far more complicated. We restrict ourselves to the statement that the Hessian matrix can be computed in quadratic time and that there is a constant number of operations that have to be performed in each iteration. Similarly to the FAST algorithm, non-maximum suppression that selects the final interest points is performed in each octave at the end of the process. This results in the following:

$$
\begin{aligned}
f_{\text{FH}}(n) \quad \in \quad & O(1) + O(1) \cdot O(n^2) \\
& + 4 \cdot \left[O(1) \cdot O(n^2) + O(1) \cdot O(n^2) \right] \\
\approx \quad & O(n^2).
\end{aligned}
\qquad (32)
$$

We – again – used the property of big O and ignored all multiplicative constants to get the final result.

Finally, let us analyze the Harris-Laplace detector. The following reasoning is based on [17] where the authors of the Harris-Laplace detector give a short analysis of computational complexity of their algorithms. The algorithm requires computation of the second order matrix consisting of image derivatives for each image point. This can be achieved by convolving the image with a Gaussian kernel which takes $O(m \cdot n^2)$, where m denotes the width of the kernel. The worst case here is that the size of the kernel will be the same as the size of the image. In that case the big O notation would be $O(n^3)$. In practice, nevertheless, the smoothing with Gaussian filter can be implemented recursively which will reduce the complexity dramatically. The candidate points are then selected and further processed. The selection process is again non-maximum suppression. "The simplified Harris-Laplace approach requires 3 convolutions of a point neighborhood with a 2D Laplacian kernel to select the scale" [17]. This takes $3 \cdot O(m \cdot p^2)$, where m denotes the size of the Laplacian kernel and p the size of the neighborhood. Since the size of the Laplacian kernel and the size of the neighborhood do not change, it takes constant time. Again, the worst case is that the number of candidate points will be the same as the number of pixels in the image. The scale selection process will then take $O(n^2) \cdot O(1)$. Finally, if we put all this pieces of information together we get:

$$\begin{aligned} f_{\text{HL}}(n) &\in O(1) + O(n^3) + O(n^2) + O(n^2) \cdot O(1) \\ &\approx O(n^3). \end{aligned} \tag{33}$$

The cubic function is a dominant term here so the resulting complexity can be written as $O(n^3)$.

The times required to calculate the sharpness coefficients more or less confirm the computational complexity analysis. VR, EOG, TEN, EOL, SML and SF have the same (quadratic) complexity. It is evident that the fastest method in all spectra is EOG. Then the rest of methods can be sorted in this order: SF, SML, EOL, TEN and VR. The differences between the methods are probably due to the level of optimization that the Java compiler performed on the loops and the operations within the innermost loop.

The methods based on the interest points detection are much slower due to the more complex equations. The values of the required times are in case of FH and FAST similar. The slowest algorithm from all 9 methods is the Harris-Laplace detector which has also the worst computational complexity in terms of O. There can also be seen big variations along all sets in FAST algorithm. It is evident that this detector works much slower in case of sets where there is a lot of details or objects. In case of simple objects like sunglasses, this method works faster. Methods like EOG, SF, SML, EOL, TEN and VR are not dependent on the content of the image and thus require nearly the same time in all sets.
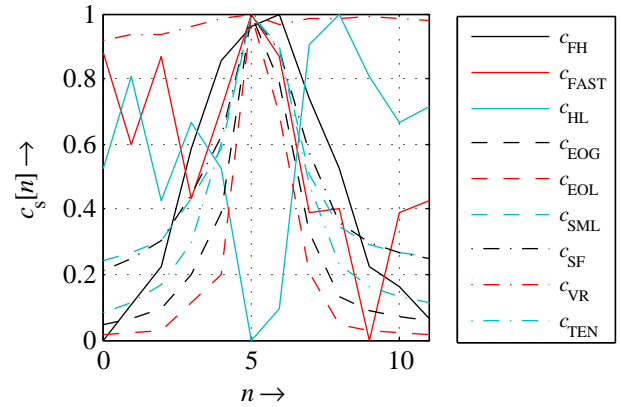

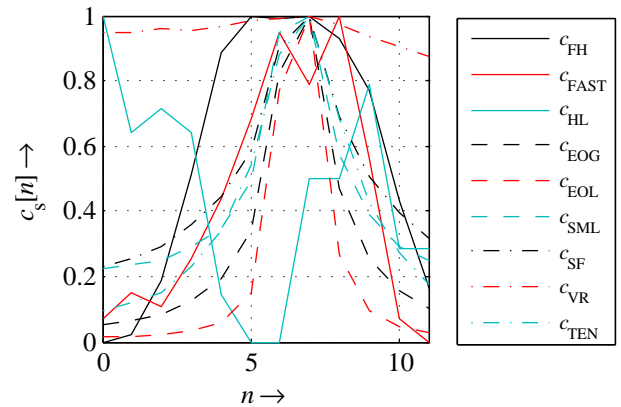**Fig. 5.** Sharpness curves for the "Guitar" (VIS) set. Sharp image for $n = 5$.

Fig. 6. Sharpness curves for the "Headphones" (VIS) set. Sharp image for $n = 7$.
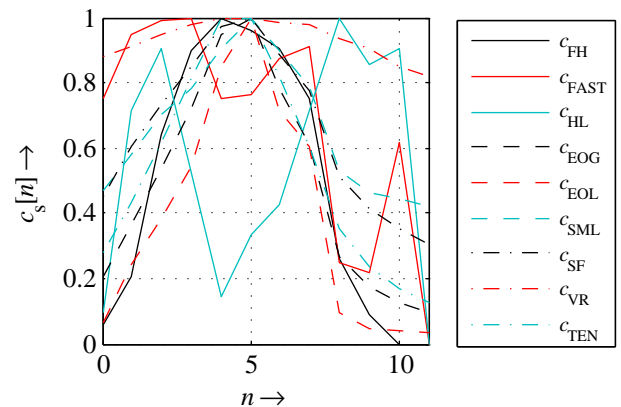

Fig. 7. Sharpness curves for the "Keyboard" (VIS) set. Sharp image for $n = 5$.

Spectrum	Object	Variables	FH	FAST	HL	EOG	EOL	SML	SF	VR	TEN	SUB
VIS	Guitar	P	6	5	8	5	5	5	5	5	5	5
		Q	0.2500	0.2500	0.2000	0.5000	1.0000	0.5000	0.5000	0.0833	0.5000	-
		mean(t) [ms]	14.9841	15.0845	227.8594	0.4665	1.3213	0.6584	0.5512	2.8794	1.9565	-
		std(t) [ms]	1.2250	1.2385	1.6132	0.0112	0.0264	0.0160	0.0133	0.0469	0.0331	-
	Headphones	P	5	5	0	7	7	7	7	7	7	7
		Q	0.1667	0.3333	0.3333	0.5000	0.5000	0.5000	0.5000	0.0833	0.5000	-
		mean(t) [ms]	14.9929	13.0836	227.2926	0.4681	1.3210	0.6593	0.5505	2.8781	1.8662	-
		std(t) [ms]	1.2399	0.6898	1.7780	0.0127	0.0248	0.0148	0.0186	0.0450	0.0346	-
	Keyboard	P	4	3	8	5	5	5	5	5	5	5
		Q	0.2000	0.1250	0.1667	0.2500	0.3333	0.2000	0.1667	0.0833	0.2500	-
		mean(t) [ms]	15.1409	13.1768	227.2032	0.4666	1.3196	0.6594	0.5514	2.8788	1.9486	-
		std(t) [ms]	1.2899	0.5501	1.4960	0.0111	0.0245	0.0163	0.0129	0.0460	0.0324	-
	Keys	P	0	2	7	2	2	2	2	1	2	2
		Q	0.1667	0.2000	0.2500	0.5000	0.5000	0.5000	0.5000	0.1429	0.5000	-
		mean(t) [ms]	14.9894	12.5209	227.4061	0.4670	1.3217	0.6605	0.5518	2.8829	1.8663	-
		std(t) [ms]	1.2776	1.0533	1.6563	0.0132	0.0255	0.0176	0.0134	0.0500	0.0394	-
	Loudspeaker	P	8	0	8	8	8	8	8	7	8	8
		Q	0.3333	0.3333	0.1429	1.0000	1.0000	1.0000	1.0000	0.0833	1.0000	-
		mean(t) [ms]	14.8682	12.9190	226.3855	0.4670	1.3197	0.6596	0.5515	2.8796	1.8336	-
		std(t) [ms]	1.2396	0.9943	1.7968	0.0126	0.0250	0.0168	0.0141	0.0496	0.0342	-
	Mixer	P	9	8	11	8	8	8	8	8	8	8
		Q	0.2500	0.5000	0.1250	1.0000	1.0000	1.0000	0.3333	0.1111	1.0000	-
		mean(t) [ms]	14.9962	16.1164	227.6758	0.4668	1.3192	0.6587	0.5503	2.8778	1.9902	-
		std(t) [ms]	1.2228	2.1885	1.5076	0.0104	0.0238	0.0152	0.0127	0.0456	0.0342	-
	Sunglasses	P	5	6	7	5	5	5	5	5	5	5
		Q	0.5000	0.2500	0.3333	0.3333	0.5000	0.2500	0.3333	0.0909	0.3333	-
		mean(t) [ms]	14.9145	8.9840	227.1235	0.4669	1.3201	0.6592	0.5507	2.8774	1.8158	-
		std(t) [ms]	1.2504	0.3864	1.4545	0.0109	0.0241	0.0151	0.0112	0.0486	0.0365	-
NIR	Building	P	6	6	20	6	6	6	6	6	6	6
		Q	0.0667	0.5000	0.2500	0.3333	0.5000	0.2500	0.1667	0.0476	0.1250	-
		mean(t) [ms]	15.1108	19.4225	222.0125	0.4774	1.3294	0.6672	0.5567	2.8817	2.0047	-
		std(t) [ms]	2.4673	2.7906	5.0679	0.1234	0.2602	0.2205	0.1242	0.0689	0.2707	-
	Car	P	6	4	20	4	4	4	4	5	4	4
		Q	0.0667	0.2500	0.2500	0.5000	0.5000	0.2500	0.2500	0.0476	0.1667	-
		mean(t) [ms]	15.0433	19.2631	227.9791	0.4680	1.3214	0.6596	0.5503	2.8782	2.0158	-
		std(t) [ms]	1.2274	1.3836	1.7366	0.0124	0.0260	0.0164	0.0111	0.0454	0.0343	-
	Corridor	P	7	4	20	7	7	7	7	7	7	7
		Q	0.1000	0.1000	0.5000	0.2500	0.3333	0.1429	0.1429	0.0476	0.1250	-
		mean(t) [ms]	14.9935	16.5919	227.5685	0.4673	1.3203	0.6584	0.5500	2.8765	1.9796	-
		std(t) [ms]	1.2361	0.5791	1.5022	0.0114	0.0234	0.0162	0.0127	0.0438	0.0328	-
	Head	P	4	19	2	4	4	4	4	4	4	4
		Q	0.3333	0.1667	0.2000	1.0000	1.0000	1.0000	1.0000	0.0667	1.0000	-
		mean(t) [ms]	14.8578	13.2972	227.0702	0.4663	1.3210	0.6585	0.5506	2.8778	1.9595	-
		std(t) [ms]	1.2591	0.7091	1.6785	0.0127	0.0237	0.0168	0.0111	0.0441	0.0328	-
	Keyboard	P	4	16	0	3	4	3	3	2	3	3
		Q	0.2500	0.1000	0.1250	0.5000	0.5000	0.0833	0.3333	0.1000	0.5000	-
		mean(t) [ms]	14.7773	11.2346	227.7390	0.4668	1.3212	0.6594	0.5514	2.8784	1.9464	-
		std(t) [ms]	1.2184	1.6294	1.5942	0.0115	0.0243	0.0171	0.0135	0.0448	0.0360	-
	Office Desk	P	6	7	10	6	6	6	6	5	6	6
		Q	0.2000	0.3333	0.0909	1.0000	1.0000	1.0000	1.0000	0.0556	1.0000	-
		mean(t) [ms]	14.9826	16.7514	227.6645	0.4667	1.3211	0.6589	0.5511	2.8777	2.0001	-
		std(t) [ms]	1.2198	1.3530	1.5088	0.0102	0.0243	0.0159	0.0118	0.0434	0.0343	-
	Pens	P	7	18	3	7	7	7	7	7	7	7
		Q	0.2500	0.2500	0.1429	1.0000	1.0000	1.0000	1.0000	0.1111	1.0000	-
		mean(t) [ms]	14.8125	13.6770	220.2136	0.4731	1.3229	0.6606	0.5515	2.8777	1.9705	-
		std(t) [ms]	1.2235	0.7047	1.5224	0.0117	0.0229	0.0161	0.0123	0.0445	0.0333	-
TH	Circuit Breakers	P	17	20	18	17	17	17	17	8	17	17
		Q	0.5000	0.0769	0.2500	1.0000	0.5000	0.0400	0.0476	0.0370	1.0000	-
		mean(t) [ms]	14.7161	22.5668	226.6245	0.4679	1.3020	0.6617	0.5529	2.8737	2.0141	-
		std(t) [ms]	2.1759	2.8530	4.3977	0.1078	0.2260	0.1924	0.1072	0.0601	0.2284	-
	Building	P	25	3	23	25	12	12	25	14	25	25
		Q	0.5000	0.0500	0.2000	1.0000	0.1000	0.0476	1.0000	0.0435	1.0000	-
		mean(t) [ms]	14.7252	24.9486	226.1934	0.4644	1.2955	0.6577	0.5496	2.8672	2.0132	-
		std(t) [ms]	1.2205	1.3633	1.9326	0.0109	0.0241	0.0158	0.0110	0.0414	0.0332	-
	Circuit	P	6	26	5	5	4	26	5	26	5	5
		Q	0.3333	0.0909	0.1250	0.3333	0.3333	0.0370	0.3333	0.0769	0.3333	-
		mean(t) [ms]	14.8656	18.8039	225.2773	0.4642	1.2956	0.6575	0.5496	2.8729	1.9697	-
		std(t) [ms]	1.2771	1.0342	1.7693	0.0100	0.0241	0.0158	0.0118	0.0433	0.0347	-
	Engine	P	15	0	16	17	14	14	17	25	17	14
		Q	0.5000	0.2000	0.2000	0.2000	0.3333	0.0370	0.1429	0.0526	0.2000	-
		mean(t) [ms]	14.8706	15.9281	225.0788	0.4645	1.2956	0.6577	0.5501	2.8706	1.9627	-
		std(t) [ms]	1.2816	0.6803	1.5495	0.0112	0.0226	0.0146	0.0124	0.0428	0.0310	-
	Printer	P	17	0	16	0	0	0	0	5	18	18
		Q	0.3333	0.0909	0.2000	0.2500	0.3333	0.0435	0.0500	0.0435	0.2000	-
		mean(t) [ms]	14.7934	20.9188	225.8288	0.4640	1.2950	0.6573	0.5494	2.8714	1.9937	-
		std(t) [ms]	1.2652	0.8253	1.5749	0.0100	0.0248	0.0147	0.0105	0.0425	0.0328	-
	Server	P	21	3	18	20	20	6	20	10	20	20
		Q	0.3333	0.1000	0.1667	1.0000	1.0000	0.0435	1.0000	0.0400	1.0000	-
		mean(t) [ms]	14.7988	21.7984	226.1382	0.4650	1.2949	0.6571	0.5488	2.8709	2.0075	-
		std(t) [ms]	1.2603	1.2813	1.6296	0.0108	0.0229	0.0161	0.0112	0.0416	0.0314	-
	Tube	P	19	0	20	0	0	2	0	2	20	20
		Q	0.5000	0.0625	0.1429	0.1429	0.1667	0.0714	0.0526	0.0667	0.2500	-
		mean(t) [ms]	14.7007	20.8095	225.3168	0.4638	1.2952	0.6565	0.5496	2.8721	1.9904	-
		std(t) [ms]	1.2266	1.6547	1.6777	0.0105	0.0226	0.0153	0.0108	0.0436	0.0335	-

Tab. 2. Comparison of measurements according to the quality factor Q, selected position of sharp image P and mean and standard deviation of required time to calculate the sharpness coefficient. FH – Fast Hessian, FAST – Features from Accelerated Segment Test, HL – Harris-Laplace, EOG – Energy of Image Gradient, EOL – Energy of Laplace Operator, SML – Sum-modified Laplacian, SF – Spatial Frequency, VR – Variance, SUB – Subjectively Selected Sharp Image.
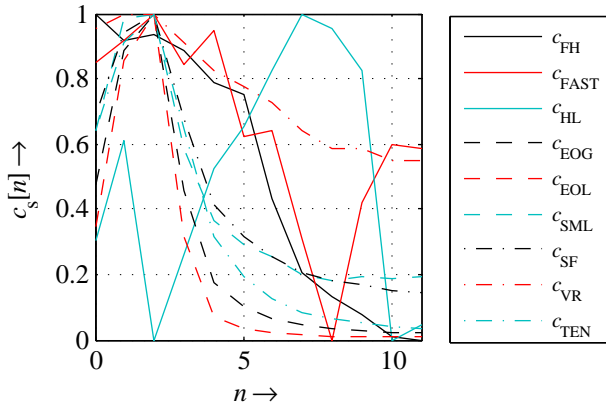
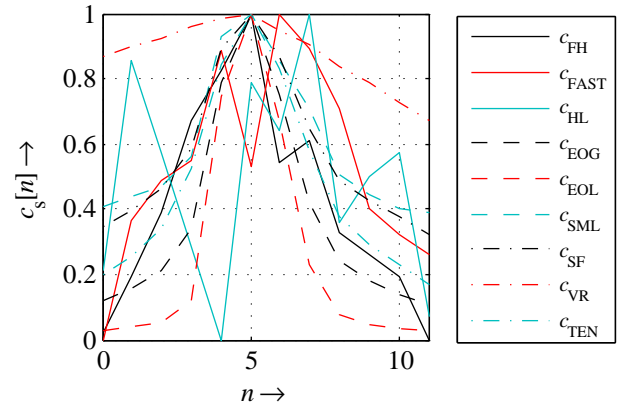**Fig. 8.** Sharpness curves for the "Keys" (VIS) set. Sharp image for $n = 2$.

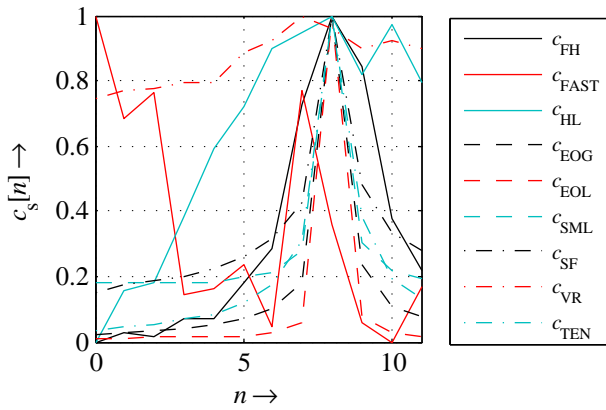**Fig. 11.** Sharpness curves for the "Sunglasses" (VIS) set. Sharp image for $n = 5$.

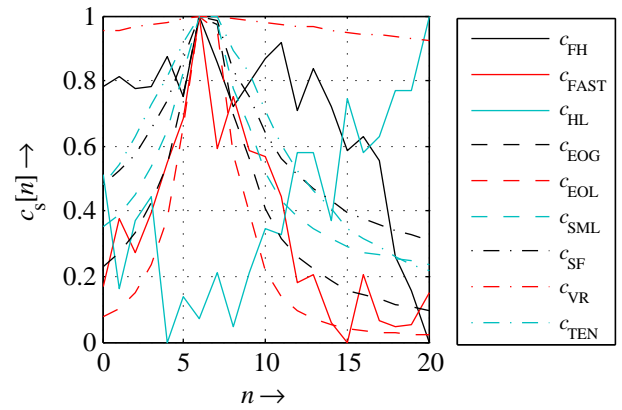**Fig. 9.** Sharpness curves for the "Loudspeaker" (VIS) set. Sharp image for $n = 8$.

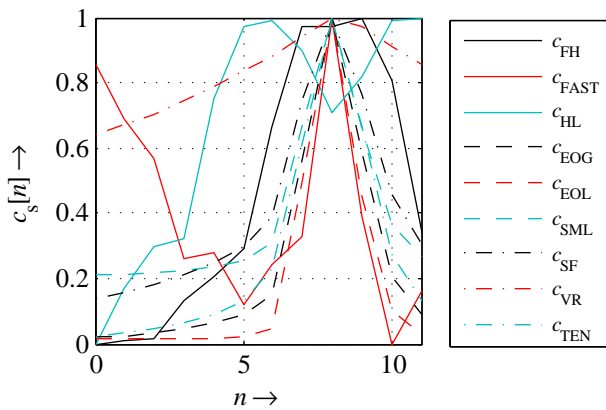**Fig. 12.** Sharpness curves for the "Building" (NIR) set. Sharp image for $n = 6$.

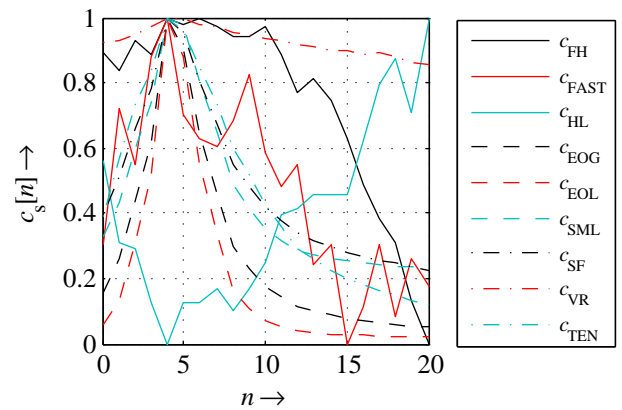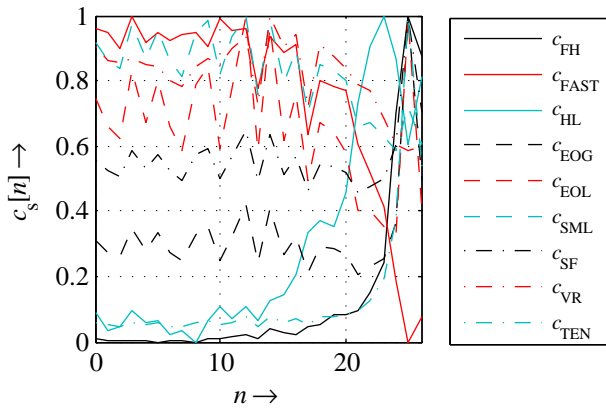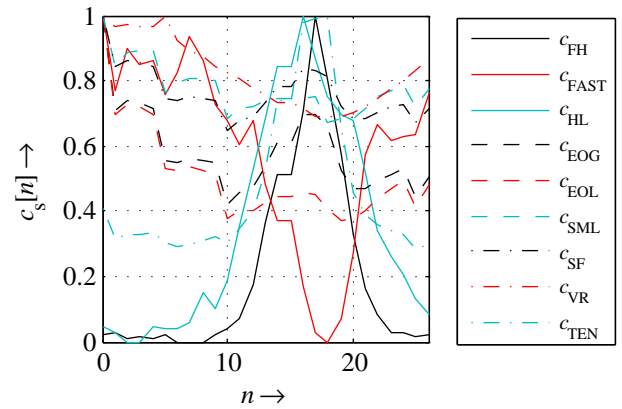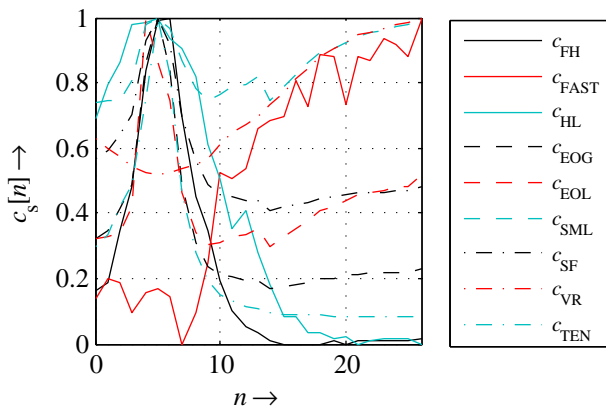**Fig. 10.** Sharpness curves for the "Mixer" (VIS) set. Sharp image for $n = 8$.

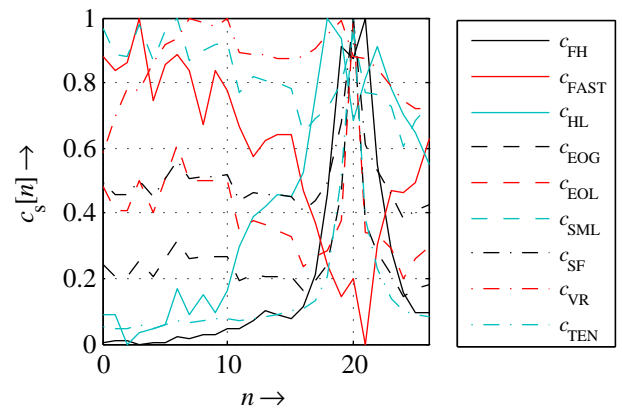**Fig. 13.** Sharpness curves for the "Car" (NIR) set. Sharp image for $n = 4$.
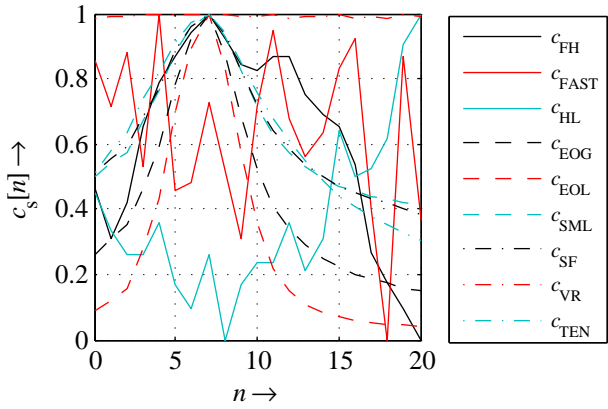
Fig. 14. Sharpness curves for the "Corridor" (NIR) set. Sharp image for $n = 7$.
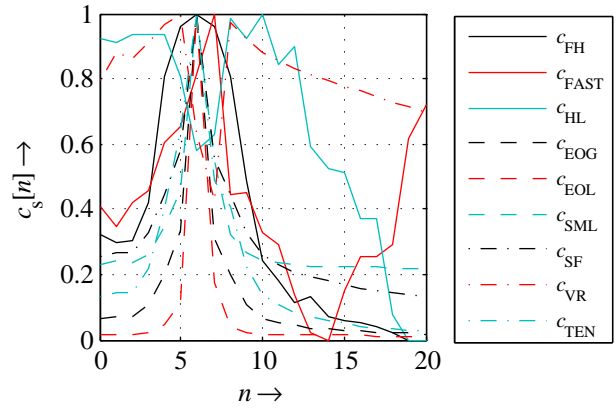
Fig. 17. Sharpness curves for the "Office Desk" (NIR) set. Sharp image for $n = 6$.
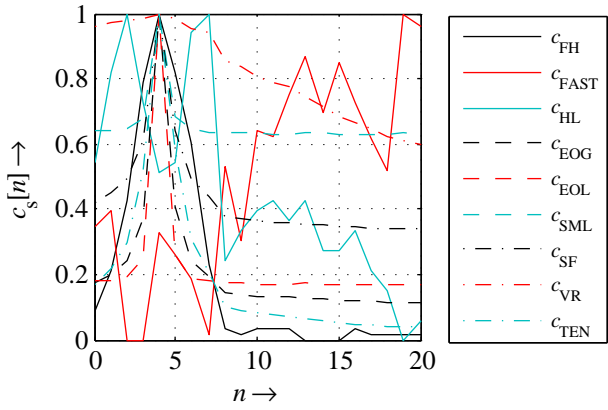
Fig. 15. Sharpness curves for the "Head" (NIR) set. Sharp image for $n = 4$.
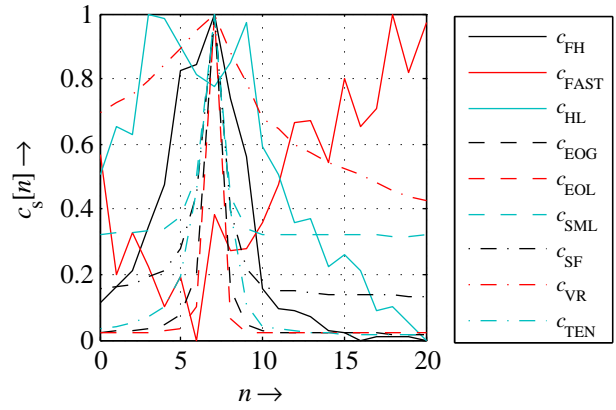
Fig. 18. Sharpness curves for the "Pens" (NIR) set. Sharp image for $n = 7$.
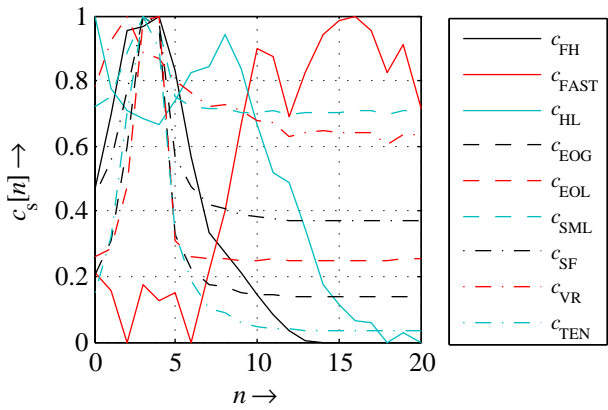
Fig. 16. Sharpness curves for the "Keyboard" (NIR) set. Sharp image for $n = 3$.
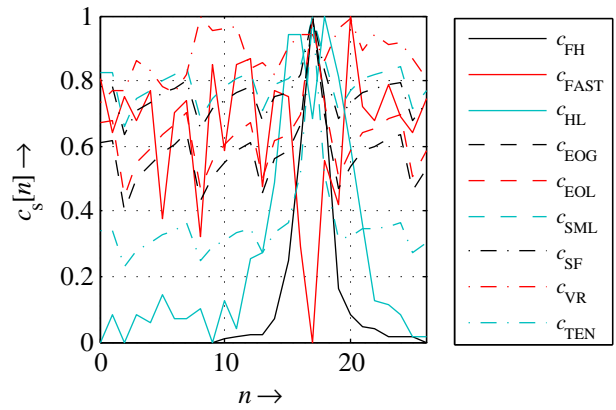
Fig. 19. Sharpness curves for the "Circuit Breaker" (TH) set. Sharp image for $n = 17$.

Fig. 20. Sharpness curves for the "Building" (TH) set. Sharp image for $n = 25$.

Fig. 21. Sharpness curves for the "Circuit" (TH) set. Sharp image for $n = 5$.

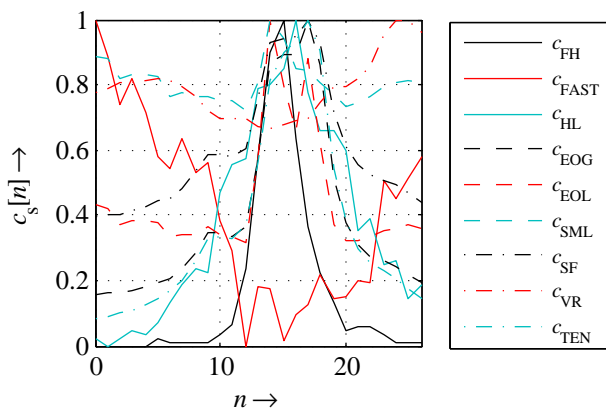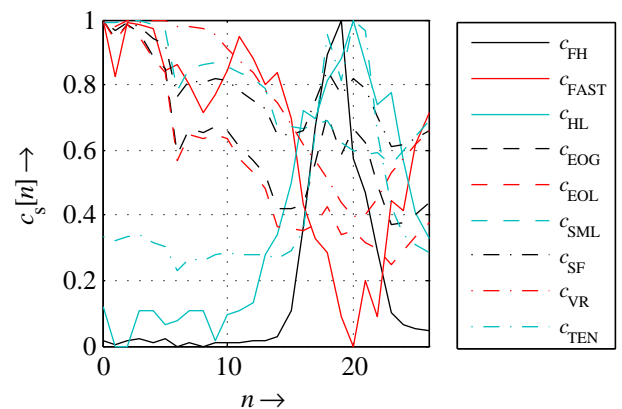**Fig. 22.** Sharpness curves for the "Engine" (TH) set. Sharp image for $n = 14$.

Fig. 23. Sharpness curves for the "Printer" (TH) set. Sharp image for $n = 18$.

Fig. 24. Sharpness curves for the "Server" (TH) set. Sharp image for $n = 20$.

Fig. 25. Sharpness curves for the "Tube" (TH) set. Sharp image for $n = 20$.

6. Conclusion

We presented and evaluated a novel focus measure in this paper. This measure is based on the number of interest points that are detected in the image by an interest point detector. The FAST, Fast Hessian and Harris-Laplace detectors were used in the experiments. We have previously proved [20] that a focus measure based on the number of detected interest points meets the requirements for a suitable focus measure in images acquired in visual spectrum. This was confirmed in this study and further extended to near infrared and thermal spectrum. It turned out that the method based on the Fast Hessian detector worked especially well on images acquired in thermal spectrum and outperformed previously published methods in terms of the shape of the sharpness curve. The results achieved by the novel method in other spectra are not so good but it has previously been shown that this method can give very good results if the process is carefully designed.

We also evaluated the focus measures in terms of computational complexity. A detailed discussion of the results can be found in Section 5. The regular methods outperformed the proposed method in terms of speed. Nonetheless, the Fast Hessian algorithm which achieved best results from the three interest point detectors can be further tuned to operate more quickly by reducing the number of scales and octaves that are processed which might result in lowering the times needed to compute the values. The Fast Hessian algorithm needed about 15 ms to process one image. This is still very good result and this algorithm can be used in applications where speed is not such an issue. Since the Fast Hessian detector gave best results in thermal spectrum, one possible application could be a multi-focus fusion system for thermal images. Such application could be of great use in military, industry as well as healthcare. This is another topic for our future work.

An additional contribution of this paper is a new database of images acquired in visual, near infrared and thermal spectrum. This database was made publicly available for free to the scientific community.

Finally, it can be concluded that the nature of the algorithm that is used to detect the interest points is especially important. The best results provided the Fast Hessian algorithm which is a blob detector based on the Hessian matrix. The other two interest point algorithms detect corners and the results suggest that the corner detectors are not well suited for measuring focus. An interesting future direction to explore might be to test other well-known interest point detectors such as the Harris corner detector and Difference of Gaussians detector [22].

Acknowledgements

References

[1] HUANG, W., JING, Z. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 2007, vol. 28, no. 4, p. 493 - 500.

[2] ESPINOSA-DURÓ, V., FAUNDEZ-ZANUY, M., MEKYSKA, J. Beyond cognitive signals. *Cognitive Computation*, 2011, vol. 3, no. 2, p. 374 - 381.

[3] ESPINOSA-DURÓ, V., FAUNDEZ-ZANUY, M., MEKYSKA, J., MONTE-MORENO, E. A criterion for analysis of different sensor combinations with an application to face biometrics. *Cognitive Computation*, 2010, vol. 2, no. 3, p. 135 - 141.

[4] ESPINOSA-DURÓ, V., FAUNDEZ-ZANUY, M., MEKYSKA, J. A new face database simultaneously acquired in visible, near-infrared and thermal spectrums. *Cognitive Computation*, 2012, vol. 5, no. 1, p. 1 - 17.

[5] FAUNDEZ-ZANUY, M., MEKYSKA, J., ESPINOSA-DURÓ, V. On the focusing of thermal images. *Pattern Recognition Letters*, 2011, vol. 32, no. 11, p. 1548 - 1557.

[6] BENES, R., DVORAK, P., FAUNDEZ-ZANUY, M., ESPINOSA-DURÓ, V., MEKYSKA, J. Multi-focus thermal image fusion. *Pattern Recognition Letters*, 2012, vol. 34, no. 5, p. 536 - 544.

[7] ZUKAL, M., CIKA, P., BURGET, R. Evaluation of interest point detectors for scenes with changing lightening conditions. In *Proceedings of the 34th International Conference on Telecommunications and Signal Processing (TSP)*. Budapest (Hungary), 2011, p. 579 - 583.

[8] KROTKOV, E. Focusing. *International Journal of Computer Vision*, 1987, vol. 1, p. 223 - 237.

[9] SUBBARAO, M., CHOI, T., NIKZAD, A. Focusing techniques. *Optical Engineering*, 1993, vol. 32, no. 11, p. 2824 - 2836.

[10] RUSS, J. C. *Image Processing Handbook,*, 4th ed. Boca Raton (FL, USA): CRC Press, 2006.

[11] NAYAR, S., NAKAGAWA, Y. Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, vol. 16, no. 8, p. 824 - 831.

[12] ESKICIOGLU, A., FISHER, P. Image quality measures and their performance. *IEEE Transactions on Communications*, 1995, vol. 43, no. 12, p. 2959 - 2965.

[13] TUYTELAARS, T., MIKOLAJCZYK, K. *Local Invariant Feature Detectors: A Survey*. Hanover (MA, USA): Now Publishers, 2008.

[14] ROSTEN, E., PORTER, R., DRUMMOND, T. Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, vol. 32, no. 1, p. 105 - 119.

[15] BAY, H., TUYTELAARS, T., GOOL, L. V. Surf: Speeded up robust features. In *Proceedings of the 9th European Conference on Computer Vision*. Graz (Austria), 2006, p. 404 - 417.

[16] VIOLA, P., JONES, M. Robust real-time object detection. *International Journal of Computer Vision*, 2001.

[17] MIKOLAJCZYK, K., SCHMID, C. Scale & affine invariant interest point detectors. *International Journal of Computer Vision*, 2004, vol. 60, p. 63 - 86.

[18] HARRIS, C., STEPHENS, M. A combined corner and edge detection. In *Proceedings of the 4th Alvey Vision Conference*. Manchester (United Kingdom), 1988, p. 147 - 151.

[19] FLIR, United States. *FLIR SC660, The High Performance infrared inspection system (datasheet)*. 2 pages. [Online] Cited 2013-03-04. Available at: http://webs.uvigo.es/ramiro/sc660datasheet.pdf.

[20] MEKYSKA, J., ZUKAL, M., CIKA, P., SMEKAL, Z. Interest points as a focus measure. In *Proceedings of the 35th International Conference on Telecommunications and Signal Processing (TSP)* Prague (Czech Republic), 2012, p. 774 - 778.

[21] JANE, O., GÖKHAN, H. A quantitative study on optimum parameters selection in adaptive unsharp masking technique for infrared images. *Radioengineering*, 2009, vol. 18, no. 4, p. 611 - 617.

[22] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, vol. 60, no. 2, p. 91 - 110.

About Authors. . .

Martin ZUKAL was born in Brno, Czech Republic. He received his M.Sc. from Brno University of Technology in 2010. He is currently Ph.D. Candidate at The Faculty of Electrical Engineering and Communication, Brno University of Technology. He participates in research & development projects as well as pedagogic activities. His research has been focused on digital image processing. His specialization covers feature extraction algorithms with an emphasis on interest point detection and evaluation. He is a member of the SPLab (http://www.splab.cz) research group that deals with digital signal processing.

Jiri MEKYSKA was born in Holesov, Czech Republic. He received the Ms.C. degree from Telecommunications in 2010 at BUT (Brno University of Technology). Currently he is a Ph.D. student, junior researcher and member of SPLab (Signal Processing Laboratory) at the Department of Telecommunications at BUT. His research interest lies in the field of speech signal analysis and its application in biomedicine. He deals with the hand-written text analysis, thermal image processing and information theory as well.

Petr CIKA was born in Brno, Czech Republic in 1982. He received the M.Sc. and Ph.D. degrees in Electronics and Communications program from Brno University of Technology in 2005, respectively 2009. He is currently an assistant at the Department of Telecommunications, Brno University of Technology. His research interests include the image and video processing and multimedia data watermarking.

Zdenek SMEKAL is a full professor at the Faculty of Electrical Engineering and Communications, Brno University of Technology. He received a professor degree in 2000 in the field of Theoretical Electrical Engineering. Since 1998 he is a member of IEEE (Signal Processing, Communications and Computer Societies), since 2006 as a Senior Member. He is also a member of Audio Engineering Society. He is a member of scientific councils (FEEC in BUT Brno and FEE in CTU Prague). He is an author of three monographs and more than 150 papers for international conferences and journals.