A biologically inspired scale-space for illumination invariant feature detection

Vasillios Vonikakis[1], **Dimitrios Chrysostomou**[2], **Rigas Kouskouridas**[2] **and Antonios Gasteratos**[2]

[1] Advanced Digital Sciences Center (ADSC), 1 Fusionopolis Way, #08-10 Connexis North Tower, Singapore 138632

[2] Laboratory of Robotics and Automation, Department of Production and Management Engineering, Democritus University of Thrace, Building I, Vasilissis Sophias 12, GR-671 00 Xanthi, Greece

E-mail: bbonik@adsc.com.sg, dchrisos@pme.duth.gr, rkouskou@pme.duth.gr, agaster@pme.duth.gr

Abstract. This paper presents a new illumination invariant operator, combining the non-linear characteristics of biological center-surround cells with the classic *Difference of Gaussians* operator. It specifically targets the underexposed image regions, exhibiting increased sensitivity to low contrast, while not affecting performance in the correctly exposed ones. The proposed operator can be used to create a scale-space, which in turn, can be a part of a *SIFT*-based detector module. The main advantage of this illumination invariant scale-space is that, using just one global threshold, keypoints can be detected both in the dark and the bright image regions. In order to evaluate the degree of illumination invariance that the proposed, as well as, other existing operators exhibit, a new benchmark dataset is introduced. It features a greater variety of imaging conditions, compared to existing databases, containing real scenes under various degrees and combinations of uniform and non-uniform illumination. Experimental results show that the proposed detector extracts greater number of features, with high level of repeatability, compared to other approaches, for both uniform and non-uniform illumination. This, along with its simple implementation, renders the proposed feature detector particularly appropriate for outdoor vision systems, working in environments under non-controlled illumination conditions.

Keywords: feature detector; biologically inspired; scale-space pyramid; illumination invariant

Submitted to: *Meas. Sci. Technol.*

1. Introduction

Difference of Gaussians (DoG) is a well-established operator in the field of computer vision, used for the extraction of edges [1] or features, as part of the Laplacian pyramid [2]. The Laplacian pyramid is part of the Scale-Invariant Feature Transform *(SIFT)* detector [3, 4], which is extensively used in many computer vision tasks [5, 6, 7] Although the *SIFT* detector has been designed in such way that it exhibits some degree of illumination invariance (the local minima and maxima keypoints in the scale-space are invariant to contrast magnitude and thus, invariant to illumination changes), non-uniform illumination conditions can still be a challenge. This is clearly depicted in Figure 1, in which, a scene is captured under three different kinds of illumination, uniform bright, uniform dim and non-uniform. For each of these three cases, the extracted keypoints and their sum total is shown, for different threshold values. As expected, in all three cases the number of extracted keypoints is inversely related to the threshold value. Furthermore, lower threshold values (cases D and E) result to the extraction of keypoints corresponding to noise and not to any surface properties. Ideally, the total number and locations of all extracted keypoints should be identical in all three images, since they depict exactly the same scene. However, there are important differences between the three types of illumination, and especially between the uniformly well exposed image and the image under non-uniform illumination.

The differences are both in the location of the extracted keypoints and in their total sum. In the case of the uniformly well-exposed image, high threshold values (cases A and B) result in the extraction of keypoints in all the regions of the foreground. On the contrary, in the case the image is captured under non-uniform illumination, the extracted keypoints are located only at the bright regions of the foreground. No keypoints are extracted on the dark image regions. Furthermore, the number of keypoints in the case of non-uniform illumination is less than half, compared to the uniformly well-exposed image. In order to extract keypoints in the dark image regions, for the case of non-uniform illumination, threshold must be set to 25% (case C) of its original value (case A). Still, in this case, the number of keypoints located in the shadows is way less than in the well-exposed image. Any attempt to decrease the threshold value even further (cases D and E), results to the extraction of keypoints not corresponding to any surface properties but to noise. Consequently, almost the whole image is covered by keypoints. The case of dim uniform illumination exhibits an intermediate state between the two extremes of bright uniform and non-uniform illumination. More specifically, for threshold cases A and B, the number of extracted keypoints, as well as their locations, are similar to the bright uniform illumination. This is in accordance with the fact that the local minima and maxima in the scale-space are invariant to the magnitude of contrast. However, as threshold values lower (cases D and E), the number and location of keypoints resembles the case of non-uniform illumination.

A similar example is shown in Figure 2, where a scene with two color checkers, under non-uniform illumination is depicted (Figure 2(a)), with one located within a

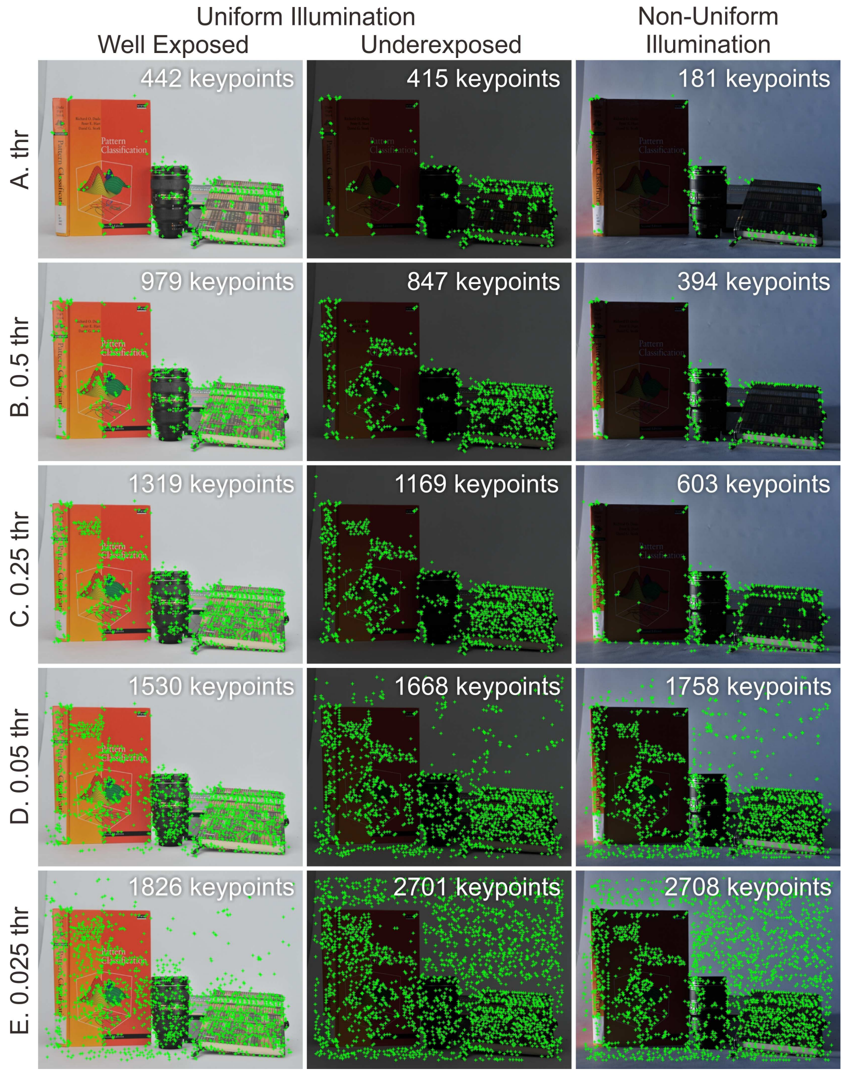
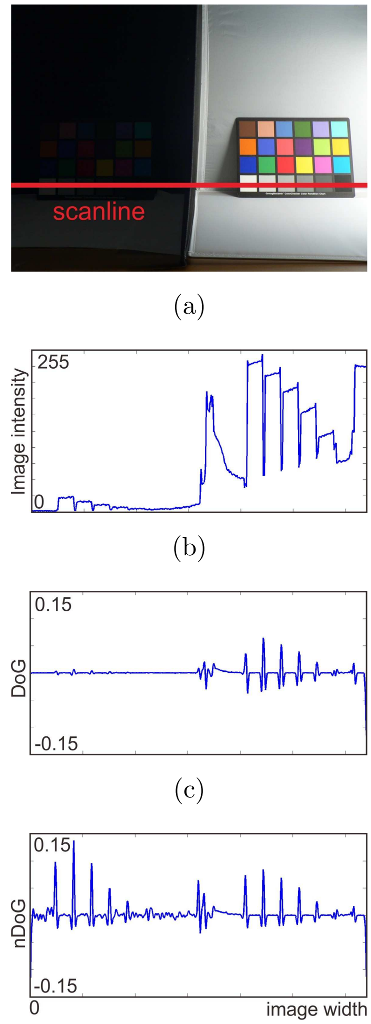


Figure 1. The extracted *SIFT* keypoints and their total number, for various threshold values, in a scene captured under three different kinds of illumination.

strong shadow and the other in a well-exposed image region. This image is part of the High Dynamic Range work-shop presented during the last CREATE (Color Research for European Advanced Technology Employment) meeting [8]. Figure 2(b) depicts a single scanline of this image, which crosses the achromatic set of boxes, for both color checkers, while Figure 2(c) depicts the output of the *DoG* operator for this specific scanline. It is evident that the magnitude of gradient in the dark image region is significantly lower than the one in the well-exposed region. In these cases, although the local extrema of the gradient will be detected both in the dark and bright regions, it is difficult to find a single global threshold that will result in the selection of keypoints in the whole image. More importantly, since this threshold has to be set quite low, in order to detect gradients of low magnitude, it may result in the extraction of keypoints that correspond to noise. The above examples demonstrate the limitations of the classic scale space regarding illumination invariance. This can have a negative impact to vision systems that operate



(d)

Figure 2. (a) A scene with two color checkers, under non-uniform illumination; (b) A single scanline of the image scene, which crosses the achromatic boxes, in both color checkers; (c) The output of the *DoG* operator for the scanline; (d) The output of the *nDoG* operator for the scanline.

under non-controlled illumination conditions. In such cases, the captured images will inevitably suffer from underexposed regions, preventing the extraction of keypoints in these areas. As a result, object recognition, or any other feature-based algorithm, will be impaired, thus, deteriorating the performance of the whole system. Consequently, any method or approach that gives a solution to this problem is of significant importance to the computer vision community.

The first attempts towards this direction introduced a new vision framework for robust object recognition in cluttered environments [4]. Existing techniques are based on appearance features holding data with local estate. Algorithms of this kind extract local features with local extend invariant to possible illumination, viewpoint, rotation and scale changes [9]. The two main sub-mechanisms of such frameworks are a detector and a descriptor of the areas of interest. The main idea underlying such mechanism is that while the interest point detector pursues points or regions in a scene containing data that are salient within their local neighborhood, the descriptor organizes the information collected from the detector in a discriminating manner, so that the image is characterized by a collection of high dimensional feature vectors. One of the first attempts for the determination of illumination-invariant features have been proposed by Westhoff et al. [10] where the quantitative bilateral symmetry of an examined scene is computed using dynamic programming and vertical symmetry images are extracted using non-maxima suppression and hysteresis thresholding. Tang et al. [11] presented a novel feature descriptor called ordinal spatial intensity distribution that provided a great degree of invariance to any monotonically increasing brightness. More recently, Yu et al. [12] examined the relationship of the relative view and illumination of the images for better image matching. In the context of illumination-invariant localization for indoor robots Lee et al. utilized a twofold approach of orthogonal lines and local descriptor-based point features [13]. Furthermore, the latest attempts in face recognition domain involved the use of Haar local binary pattern features by [14] and neighboring wavelet coefficients for great illumination invariance during the extraction of local features [15].

The contribution of this paper is twofold. First, it introduces a new *DoG*-based operator, inspired by the center-surround cells of the *Human Visual System (HVS)*, which exhibits improved illumination invariant characteristics, compared to classic *DoG*. This operator can be used for the creation of an illumination invariant scale-space, which can improve scale-space based local detectors, like *SIFT*, by increasing their robustness in various kinds of illumination changes. More specifically, the proposed scale-space exhibits improved response in the underexposed image regions and exactly the same response, with the classic *DoG*-based scale-space, for the well-exposed image regions. As a result, it ensures that a single global threshold can extract keypoints both in the shadows and in the bright areas, avoiding at the same time the extraction of those corresponding to noise. Additionally, the proposed scale-space is simple to implement and incorporate in existing *SIFT*-based vision systems, thus, enhancing their illumination invariance, especially for non-uniform illumination conditions, while not affecting their performance in bright uniform illumination. Consequently, it can boost the performance of vision systems which operate in non-controlled illumination environments.

The second contribution of this paper is a new dataset specifically targeted to evaluate the illumination invariance of vision systems. Unlike existing datasets, the proposed is the only one featuring scenes under various degrees and combinations of uniform and non-uniform illumination. As a result, to the best of our knowledge, it constitutes the only existing dataset that can provide clues on how the performance of algorithms may vary according to different illuminations and imaging conditions. The remainder of the paper is organized as follows: Section 2 briefly describes the biological background upon which the proposed method is based. Section 3 describes the proposed biologically-inspired scale space. Section 4 presents the new benchmark database. The experimental results are presented in Section 5 and concluding remarks are made in

Section 6.

2. Biological background

2.1. Biological center-surround operators

Neurophysiological studies have revealed that the receptive fields of the retinal ganglion cells, as well as those of other center-surround cells in the *HVS*, can be modeled as *DoG* operators [16]. Contrary to the classic *DoG* operator though, the center-surround cells of the HVS exhibit non-linear responses. Interestingly, the nonlinear response of ganglion cells is thought to contribute to illumination invariance and contrast enhancement [17]. According to the standard retinal model [18, 19], the output X_{ij} of an ON-center OFF-surround cell at grid position (i, j), obeying the membrane equations of physiology is given by

$$\frac{dX_{ij}(t)}{dt} = g_{leak}(X_{rest} - X_{ij}) + C_{ij}(E_{ex} - X_{ij}) + S_{ij}(E_{inh} - X_{ij}) \tag{1}$$

with

$$C_{ij} = \sum I_{pq} G_{\sigma C}(i - p, j - q) \tag{2}$$

$$S_{ij} = \sum I_{pq} G_{\sigma S}(i - p, j - q) \tag{3}$$

where g_{leak} is a decay constant and I is a luminance distribution (i.e. the image formed in the photoreceptor mosaic). X_{rest} (the cell's resting potential), E_{ex} (excitatory reversal potential) and E_{inh} (inhibitory reversal potential) are constants related to the neurophysiology of the cell. $G_{\sigma C}$ and $G_{\sigma S}$ are Gaussians representing the center and the surround of the cell's receptive field respectively, which are assumed to be normalized in order to integrate to unity. The steady-state solution of equation 1 is given by:

$$X_{ij,\infty} = \frac{C_{ij}E_{ex} + S_{ij}E_{inh}}{g_{leak} + C_{ij} + S_{ij}} \tag{4}$$

Equation 4 summarizes the difference between the nonlinear *DoG* operator in biological vision and its linear counterpart used in computer vision. When $E_{ex} = 1$ and $E_{inh} = -1$, which is usually the case for center-surround cells, the numerator of equation 4 is a standard linear *DoG* operator. However the denominator consists of a Sum of Gaussians *(SoG)* augmented by the decay constant g_{leak}. This acts as a multiplicative gain control, where, with increasing activity of both center and surround (i.e. with increasing luminance), the cell's response will decrease. On the other hand, in low luminance conditions, the cell's response is increasing, due to the low activity of center and surround in the denominator. As a result, center-surround cells in biological visual systems exhibit a normalized response, invariant to different illumination conditions. Since the Laplacian pyramid has already a biologically-plausible *DoG* architecture, equation 4 can be rewritten in a more compatible way to the classic scale-space, by utilizing the adjacent scales of the Gaussian pyramid. We call this operator *normalized Difference of Gaussians - (nDoG)*.

$$nDoG(i,j,\sigma) = \begin{cases} \dfrac{L(i,j,\kappa\sigma) - L(i,j,\sigma)}{L(i,j,\kappa\sigma) + L(i,j,\sigma)}, & \text{if } L(i,j,\kappa\sigma) + L(i,j,\sigma) \neq 0 \\ 0, & \text{else} \end{cases} \tag{5}$$

with

$L(i,j,\kappa\sigma) = G(i,j,\kappa\sigma) * I(i,j) = S_{ij}$

$L(i,j,\sigma) = G(i,j,\sigma) * I(i,j) = C_{ij}$

where I is the input image, G is the Gaussian function, L is the blurred image resulted by the convolution of I and G, (i,j) are the spatial coordinates, κ is a multiplicative factor that determines the different levels of blurring between adjacent scales and σ is the standard deviation of the Gaussian. $L(i,j,\kappa\sigma)$ and $L(i,j,\sigma)$ can be thought as the surround S_{ij} and the center C_{ij}, respectively, of a center-surround receptive field of the HVS. In the rest of the paper we will use the notation of center C, and surround S, to denote the fine $L(i,j,\sigma)$ and coarse $L(i,j,\kappa\sigma)$ adjacent scales, respectively, in a Gaussian pyramid.

2.2. Comparison between DoG and nDoG

Figure 2(d) depicts the response of the $nDoG$ operator for the scanline of Figure 2(b). The main difference between the classic DoG operator and $nDoG$ is clearly evident when comparing Figure 2(c) with Figure 2(d). More specifically, the $nDoG$ operator exhibits an increased response in the underexposed image region, by almost a factor of 15, compared to the DoG, and an almost identical response to DoG for the well-exposed region. As a result, the $nDoG$ operator is more invariant to local illumination changes. The main reason for this discrepancy between the two operators is evident in Equations 6 and 7, which define them as a function of local contrast differences $S - C$.

$$nDoG = \frac{S-C}{S+C} = \frac{S-C}{S+C-C+C} = \frac{S-C}{S-C+2C} = \frac{x}{x+2C} = \frac{x}{x+A} = f(x) \tag{6}$$

and

$$DoG = \frac{S-C}{B} = \frac{x}{B} = g(x) \tag{7}$$

with S representing the surround, C the center, B the maximum value that S or C may take and $x = S - C$ is the local contrast differences. $nDoG$ exhibits a non-linear response to x, adjusted by parameter A and described by function f. This function is a form of the Naka-Rushton function [20] which has been identified in many vision-related cell types and has been associated with the enhancement of contrast sensitivity in the HVS [21]. On the other hand, DoG has a linear response to x, described by function g. Figure 3 depicts the graph of function f, for various values of A, in comparison to function g. It is evident that for small values of A, f exhibits a steeper non-linear response. This non-linearity ensures that even low input values x, will result to high output responses

$f(x)$. On the contrary, since function g is linear, low input values x will result to low output responses $g(x)$. This essentially means that the *nDoG* operator has an increased response to lower local contrast, which is the case for underexposed image regions.

Although *nDoG* may exhibit an improved response to shadows, compared to *DoG*, it presents an important drawback that prevents its direct use for the creation of a scale-space, i.e. it does not exhibit a constant maximum output. This is clearly depicted in Figure 3, in which, *fmax* fluctuates according to the parameter A. This is more evident for high local contrast values, near the maximum value B. In practice, this essentially means that for bright image regions, *nDoG* will exhibit a lower response, compared to *DoG*. Consequently, the same threshold will result into the extraction of fewer keypoints for *nDoG*.

Figure 4 depicts the location and the number of extracted keypoints (using always the same threshold) for both *nDoG* and *DoG*, in a scene captured with different exposures. In the overexposed image, the number of extracted keypoints for *DoG* is approximately double, compared to *nDoG*. This is a direct result of the decreased output of the former in bright image regions. As exposure decreases though, so does the number of extracted keypoints for *DoG*. Consequently, in the case of the underexposed image, *DoG* results into approximately five times less keypoints, compared to the overexposed image. Contrary to *DoG*, *nDoG* exhibits the opposite behavior; as exposure decreases, the number of extracted keypoints increases. As a result, in the underexposed image, *nDoG* results into approximately five times more keypoints, compared to the overexposed one. This example demonstrates the complementary characteristics of these operators and implies that a combination of the two could result into a more robust behavior in terms of

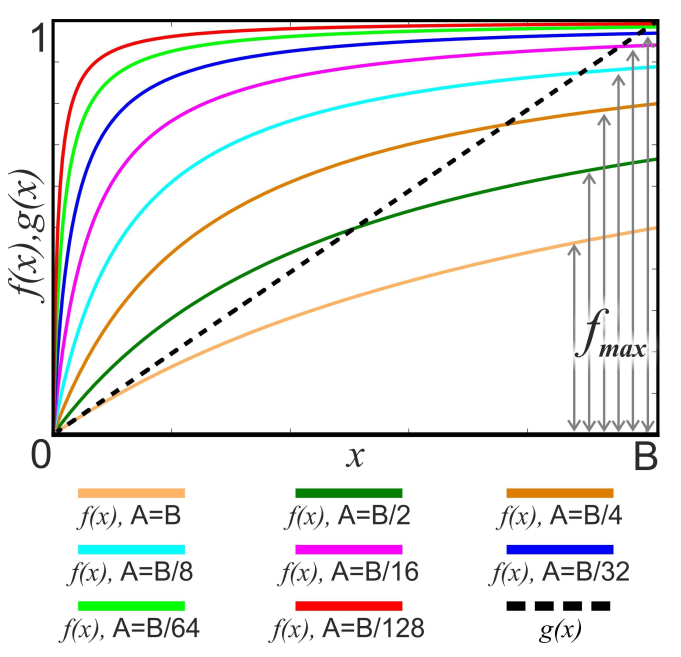


Figure 3. The graph of function f (*nDoG*), for various values of A, in comparison to function g (*DoG*).

illumination invariance.

3. Proposed operator and scale-space

According to the SIFT algorithm, a threshold is used to discard scale-space local extrema, caused by low gradient magnitude, since most of the times these points correspond to noise and not to surface properties. This approach however, may result into sacrificing the extraction of keypoints in dark image regions, and thus, impair the performance of vision systems operating in non-controlled illumination conditions. In order to avoid this unwanted behavior, the proposed method attempts to meet the following two requirements:

- Improve the response of the DoG operator in the underexposed regions, in order to extract keypoints that correspond to surface properties and not noise.
- Maintain exactly the same response with DoG in the correctly exposed and overexposed regions.

The first requirement ensures that there will be no sacrifice of extracted keypoints in shadows, while trying to avoid the extraction of noise-related features. The second requirement ensures that no performance changes will take place in existing systems

Figure 4. The location and number of extracted keypoints for $nDoG$ and DoG, in a scene captured with different exposures.

that rely on the extraction of features based on the DoG scale-space. These two requirements essentially indicate that the improvement should be specifically targeted only to underexposed regions, without affecting the already good performance of DoG in all the other parts of the image.

In order to achieve this objective, we combine DoG and $nDoG$ into one piecewise function that will selectively use one of the two operators in the appropriate cases. To further investigate the properties of the two operators and define the cases in which each one could be used, the 3-dimensional graphs of $nDoG$ and DoG are depicted in Figure 5(a) and Figure 5(b), respectively. These graphs essentially plot all the outputs for every possible combination of a center C and a surround S within the interval $[0, B]$. An apparent difference between the two graphs is when the center and the surround comprise small values near 0. This is the case of underexposed image regions, and as shown previously, $nDoG$ exhibits a strong non-linear response, compared to the linear one of DoG. Another, not so obvious difference, between the two graphs is when the center or the surround have values near B. This is the case of bright image regions, in which, DoG was found to exhibit better behavior compared to $nDoG$. In order to illustrate more clearly the dissimilarities between $nDoG$ and DoG, the 3-dimensional representation of their output differences ($nDoG - DoG$) is depicted in Figure 5(c). Additionally, Figure 5(d) depicts the center-surround plane of Figure 5(c).

From these two graphs, as well as Equations 6 and 7, it is evident that the two operators have identical outputs only when $C = S(DoG = nDoG = 0)$ and $C + S = B$ ($DoG = nDoG = (S - C)/B$). These two cases define two lines which divide the center-surround plane shown in Figure 5(d) into four quad-rants; $Q1$, $Q2$, $Q3$ and $Q4$, respectively. In every one of these quadrants, the output of one operator is always greater than the other.
$Q1$ is defined as $(C > S) \cap (C + S > B)$. In this case we have:

$$\left.\begin{matrix} S - C < 0 \\ S + C > B \end{matrix}\right\} \Rightarrow \frac{S-C}{S+C} > \frac{S-C}{B} \Rightarrow nDoG > DoG \tag{8}$$

Similarly, $Q2$ is defined as $(C > S) \cap (C + S < B)$ and in this case:

$$\left.\begin{matrix} S - C < 0 \\ S + C > B \end{matrix}\right\} \Rightarrow \frac{S-C}{S+C} < \frac{S-C}{B} \Rightarrow nDoG < DoG \tag{9}$$

$Q3$ is defined as $(C < S) \cap (C + S < B)$ and

$$\left.\begin{matrix} S - C > 0 \\ S + C < B \end{matrix}\right\} \Rightarrow \frac{S-C}{S+C} > \frac{S-C}{B} \Rightarrow nDoG > DoG \tag{10}$$

Finally, $Q4$ is defined as $(C < S) \cap (C + S > B)$ and

$$\left.\begin{matrix} S - C > 0 \\ S + C > B \end{matrix}\right\} \Rightarrow \frac{S-C}{S+C} < \frac{S-C}{B} \Rightarrow nDoG < DoG \tag{11}$$

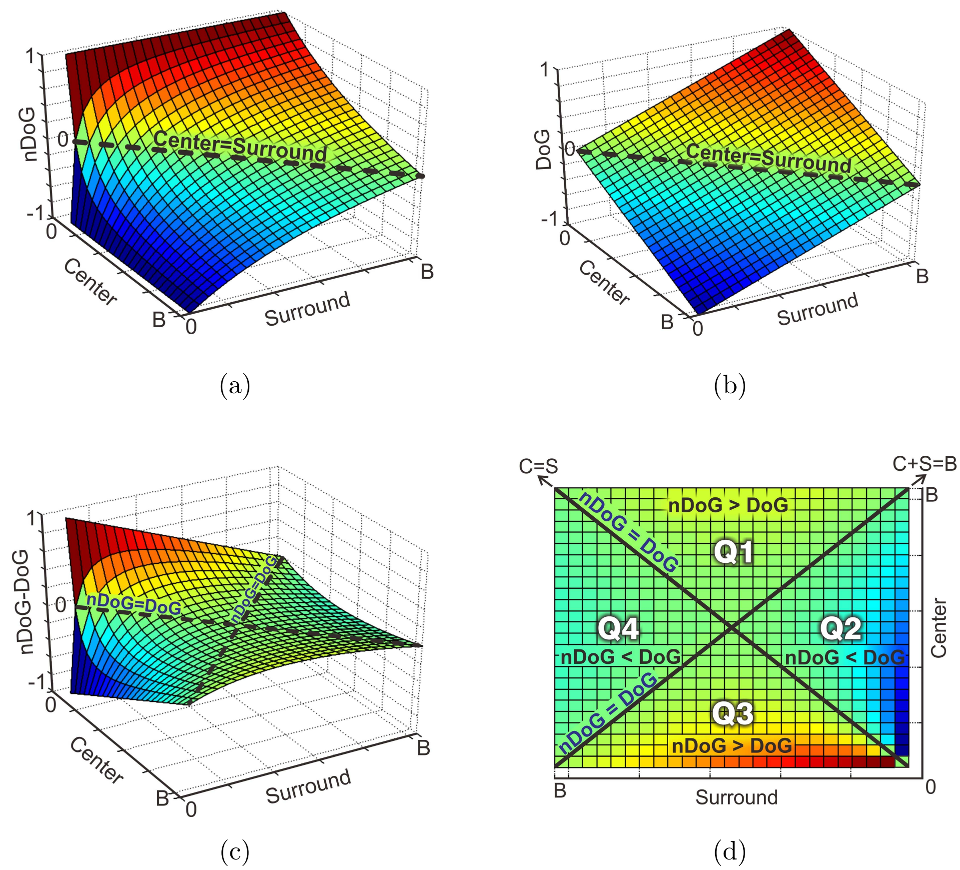


Figure 5. (a) The 3-dimensional graph of the $nDoG$ operator; (b) The 3-dimensional graph of the DoG operator; (c) The 3-dimensional graph of the difference between $nDoG - DoG$; (d) The 2-dimensional projection of the difference between $nDoG - DoG$ into the center-surround plane.

Taking into consideration the requirements mentioned above, it is obvious that we have to differentiate between dark and bright image regions. A straight forward way is to use the sum of C and S as an indicator. As it is evident from Figure 5(d), the line $C + S = B$ divides all the possible values into two sets: $Q1 \cup Q4$, in which $C + S > B$ and thus $(C + S) \in (B, 2B]$, and $Q2 \cup Q3$, in which $C + S < B$ and thus $(C + S) \in (0, B)$. Sum values in the interval $(0, B)$ can be considered to result from dark image regions, since both center and surround have low values in these regions. On the other hand, sum values in the interval $(B, 2B]$ result from bright image regions, since center and surround have higher values. Using this as an indicator for bright and dark image regions, we incorporate these requirements into the following piecewise function:

$$iiDoG = \begin{cases} nDog : \dfrac{S - C}{S + C} & \text{if } C + S < B \\ 0 & \text{if } C = S = 0 \\ DoG : \dfrac{S - C}{B} & \text{if } C + S > B \end{cases} \tag{12}$$

where *iiDoG* is the proposed *illumination invariant Difference of Gaussians* operator. Interestingly, one can reach the same result using a whole different approach for combining $nDoG$ and DoG. Since $SIFT$ uses a global threshold to discard keypoints of low gradient magnitude, it is valid to assume that selecting always the response of the operator with the higher absolute output, will usually result into the extraction of greater number of keypoints. In order to achieve this behavior, one has to select the maximum value between DoG and $nDoG$, when both are positive and the minimum value when both are negative. Since the two operators have the same numerator *(S-C)*, and their denominators are always positive non-negative values, they will always have the same sign. Additionally, the line *C-S=0* is the boundary in which the sign changes either to positive or to negative. This line divides all the possible values into two sets: $Q1 \cup Q2$, in which both $nDoG$ and DoG are negative, because $C > S$, and $Q3 \cup Q4$, where both are positive, since $S > C$. Consequently, one should select the operator with the smaller output in quadrants $Q1$ and $Q2$ ($Q1 : DoG, Q2 : nDoG$ and the operator with the greater output in quadrants $Q3$ and $Q4$ ($Q3 : nDoG, Q4 : DoG$). This is summarized in the following equation

$$iiDoG = \begin{cases} min[nDoG, DoG] & \text{if } S - C < 0 \\ 0 & \text{if } C = S = 0 \\ max[nDoG, DoG] & \text{if } S - C > 0 \end{cases} \tag{13}$$

Equation 13 can be also rewritten as:

$$iiDoG = max[[DoG]^+, [nDoG]^+] + min[[DoG]^-, [nDoG]^-] \tag{14}$$

with $[\cdot]^+ = max[\cdot, 0]$ and $[\cdot]^- = min[\cdot, 0]^-$. In particular, equation 14 is more appropriate for array-based implementations, like in Matlab, since, once the DoG and $nDoG$ output arrays have been computed, it provides the final result using simple max/min operations between them.

Equations 12, 13, 14 are all equivalent and their 3-dimensional graph is depicted in Figure 6, which essentially is a combination of Figure 5(a) and Figure 5(b). The *iiDoG* operator combines the strengths of *DoG* and *nDoG*, while avoiding at the same time their drawbacks. More specifically, *iiDoG* exhibits the illumination invariance characteristics of *nDoG*, in the underexposed image regions, while maintaining the already good performance of *DoG* in the bright image regions. Figure 7 depicts the proposed scale-space, employing the *iiDoG* operator. The main advantage of the proposed approach is that using the global threshold of *SIFT*s detector, keypoints can be extracted both in the correctly exposed image regions and in the shadows. More importantly, the improvement is strictly targeted to the underexposed image regions, with no departures from the performance of the classic *SIFT* in the bright and well exposed areas. Taking also into account the fact that the implementation of the proposed scale-space is very simple, it can be used for improving the illumination invariance of *SIFT*-based vision systems.

The proposed approach could be seen as a spatial Automatic Gain Control (AGC) method. Apart from the computer vision and image processing domain, AGC techniques

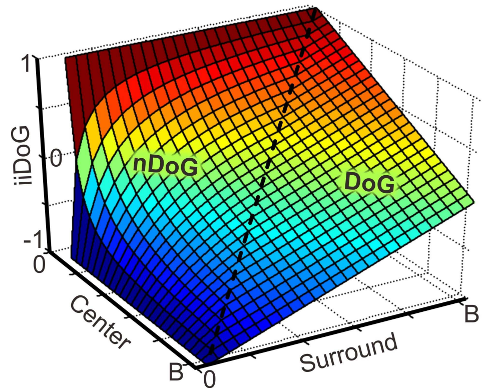


Figure 6. The 3-dimensional graph of the proposed *iiDoG* operator.

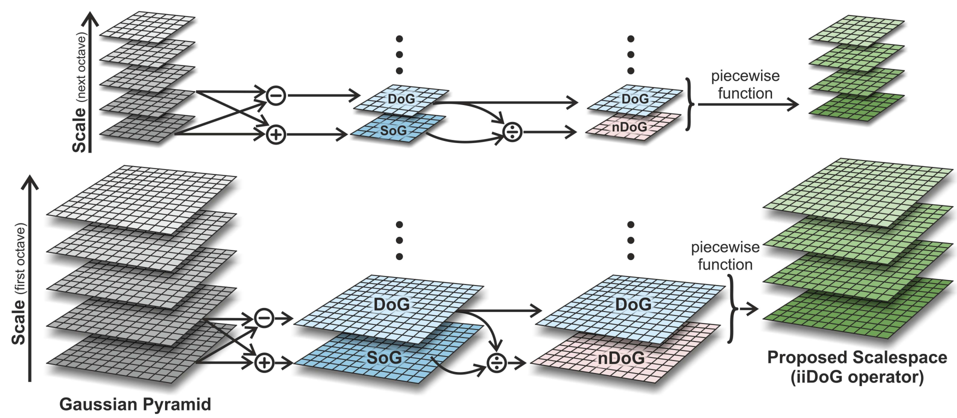


Figure 7. The proposed scale-space, based on the *iiDoG* operator.

have been proposed in other disciplines as well, such as geophysics, in order to balance different kinds of signals, e.g. aeromagnetic data. Two notable methods in this context are [22] and the *Theta map* [23], with the former, presenting better results than the latter. Figure 8 depicts a comparison between the proposed *iiDoG*, *DoG*, and *DoG+manual gain* methods along with one proposed by Cooper, for the scanline of Figure 2. For the *DoG+manual gain* method, a gain of $\times 20$ was applied only to the underexposed image region. Compared to this, the proposed method exhibits an almost equal amplification of the original *DoG* signal, in the underexposed region, while keeping it untouched in the correctly exposed. More importantly though, there is lower enhancement of noise in the underexposed region. This is not the case however with Cooper's approach. When the amplification of the underexposed region is significant ($k = 0.001, k = 0.0001$) there is also considerable enhancement of noise. As a result, this will result into the extraction of many noisy feature points by the *SIFT* detector. Additionally, the signal in the correctly exposed image region is affected, and consequently, this would change the performance of a *SIFT*-based system, if the method presented in [22] was used as an AGC. Finally, this method is based on the Hilbert transform, and as a result, every level of the Gaussian Pyramid should be transferred to the frequency domain. This inevitably would increase the computational cost. On the contrary, this is not the case for the proposed method,

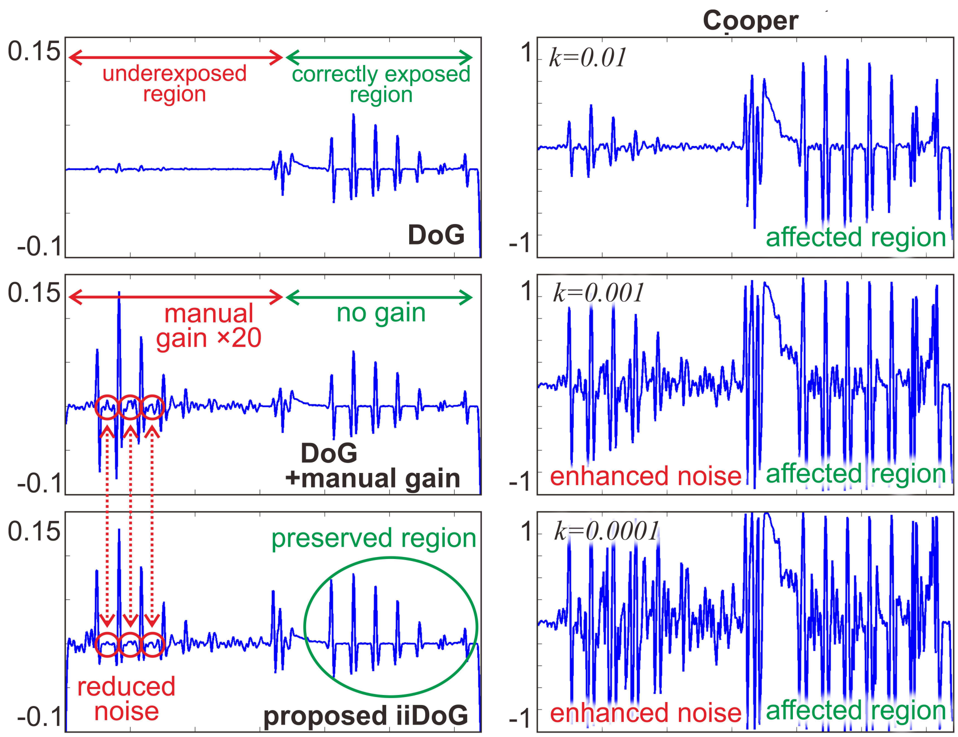


Figure 8. Comparison of the proposed *iiDoG* with automatic gain control methods.

since it is applied directly on the spatial domain.

4. Phos benchmark image database

In order to test the proposed approach, a new benchmark database has been constructed, aiming to evaluate the performance characteristics of feature detectors under various illumination conditions. The name of the proposed image database is *Phos*, which in Greek means *light*. Existing datasets focus on different viewpoints, rotation and zooming of the scenes [24], in order to test the invariance of systems in these categories. Very little attention is given, though, to the actual illumination conditions, which may exist outdoors. The vast majority of previously presented benchmarks, regarding illumination invariance, are done by manually adjusting image brightness with an image processing software. One significant exception is the *Leuven* sequence presented by Mikolajcyk and Schmid [25] where the illumination changes were occurred by adjusting the cameras aperture. This approach, however, is far from realistic. The algorithm that adjusts the brightness in an image processing software, does not necessarily exhibit the same results as the ones resulting by the exposure of a camera under real conditions.

More importantly, as the comparison of Figure 1 showed, underexposed image regions tend to have lower signal-to-noise ratio, making it difficult to distinguish between keypoints corresponding to surface properties and keypoints corresponding to noise. Consequently, taking a well exposed image, with overall good signal-to-noise ratio, and manually lowering its brightness, will not have the same effect as if the same scene would have been captured under lower illumination conditions. Furthermore, illumination in outdoor scenes is usually non-uniform. Multiple light sources, shadows and high dynamic range imagining conditions may dramatically affect the quality of captured images. As a result, any camera system functioning outdoors, will inevitably exhibit a performance reduction due to the above reasons. Undoubtedly, it is very important to measure this reduction. However, currently, there are no benchmark image databases which can be used for evaluating the performance of algorithms under more realistic lighting conditions.

The main objective of the new image database is to fill this gap in the existing benchmark databases, by specializing in realistic illumination conditions. More particularly, every one of the 15 scenes of the database contains 15 different images: 9 images captured under various strengths of uniform illumination, and 6 images under different degrees of non-uniform illumination. The images contain objects of different shape, color and texture. Moreover, the objects are positioned in random locations inside the scene. Figure 9 depicts one scene from the new image database. *Phos* database is publicly available at [26].

Uniform illumination (first row of Figure 9) is achieved using multiple diffusive light sources, evenly distributed around the objects, and a Lambertian white background. The different strengths of uniform illumination are captured by adjusting the exposure of the camera between -4 and +4 stops from the original correctly exposed image. Thus, for every scene four underexposed and four overexposed images with uniform illumination were captured. Non-uniform illumination (second row of Figure 9) is accomplished by adding a strong directional light source to the diffusive lights located around the objects. By adjusting the strength of the diffusive lights, six different mixtures of uniform and non-uniform illumination were created, ranging from both directional and uniform illumination to directional illumination only. This set of images is particularly challenging for feature detectors due to high dynamic range conditions. It contains strong shadows, which deteriorate the performance of local feature detectors. The strength of the *Phos* dataset lays in the fact that the induced shadows (uniform or non-uniform) are created incrementally. This offers the unique opportunity to study how the performance of feature detectors varies as the degree of shadows increases.

5. Experimental Results

In this section the experimental results of the performance of the proposed detector are presented and discussed. The performance of the new modified detector is compared with other widely used detectors for illumination and photometric variations in the proposed image database *Phos* and in the *Leuven* sequence presented in [25] and provided by [27].

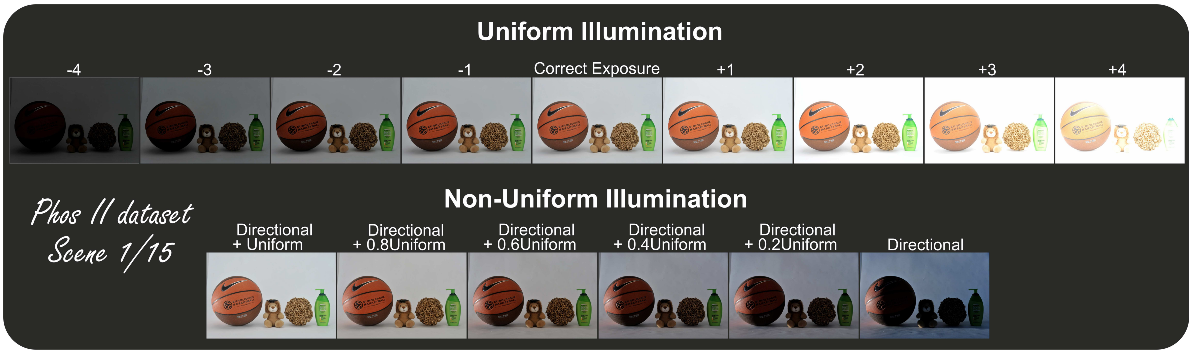


Figure 9. One scene from the proposed *Phos* dataset.

5.1. Evaluation criterion

The criterion used to evaluate a feature detector is the repeatability score the detector achieves between a given pair of images. More precisely this is the ratio between the number of region-to-region correspondences and the smaller number of regions detected in one of the images [28]. The evaluation procedure is similar to [29] which encompass only the features located in the part of the scene appearing in both images under comparison, to be taken into consideration. First, the homography between the pair of images is estimated, in order to calculate the ground truth measurement of the possible transformation. Given the estimated homography, the projected position of features and the corresponding regions of the two images are calculated and the amount of the overlap is verified. The overlap error between corresponding regions is the ratio $(1 - intersection/union)$ of the elliptic regions and it is analytically computed using the ground truth transformation. The repeatability score depends on the overlap error. Therefore, in order to be evaluated, different overlap errors are computed as well.

5.2. Test data and results

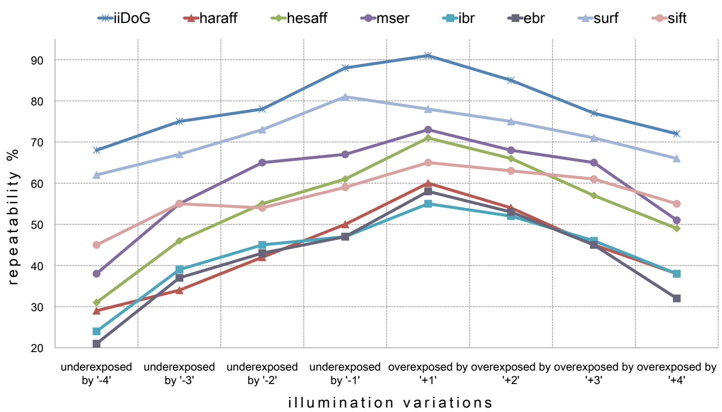
The proposed *iiDoG* operator is used for the creation of a scale-space. This scale-space is integrated in a *SIFT*-based detector, using exactly the same parameters (threshold, scales etc.) with the classic *SIFT* detector. In order to test the performance of the proposed detector three major experiments were conducted. The first one is conducted using the proposed image database, *Phos*, in order to test the illumination invariance of the proposed detector, compared to the performance of others. The algorithms used for the testing were the *Maximally Stable Extremal Region - (mser)* detector [30], the *Harris-Affine - (har-aff)* [31], the *Hessian-Affine - (hesaff)* [31], the *intensity extrema-based region detector - (ibr)* [29], the *edge-based region detector - (ebr)* [32], the original *SIFT* detector [4] and the detector module of *SURF* [9]. All these detectors were tested, along with the proposed, for repeatability, overlap error and the number of correspondences.

Figure 10 depicts the evaluation of *iiDoG* detector for the case of uniform illumination in the *Phos* dataset (first row of Figure 9). The correctly exposed image was used as reference and each of the others *(+4,+3,+2, +1, -1, -2,-3-,4)* as subjects for comparison. The results of this experiment clearly demonstrate that the proposed detector, outperforms all the other detectors in repeatability, as the exposure varies (10(a)), and when the overlap error becomes larger (10(b)). Additionally, that the proposed detector exhibits the higher number of corresponding regions in 5 out of the total 8 cases (10(c)). More importantly, in cases where the *iiDoG* is not first, it is only marginally outperformed by other detectors, ranked second among all the others.

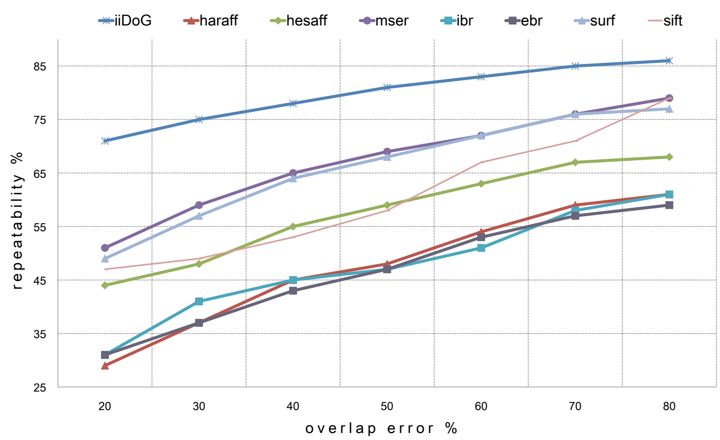
Figure 11 depicts the performance of the tested algorithms for various degrees of non-uniform illumination in the same scene (second row of Figure 9). Similarly to the case of uniform illumination, the proposed *iiDoG* operator and its resulting detector, clearly outperform all the other methods in repeatability, both when the strength of the illumination varies (11(a)) and when the overlap error becomes larger (11(b)). Additionally, the proposed detector exhibits the higher number of corresponding regions in all the test cases. More importantly, in this category, the *iiDoG* detector outperforms the second one ($SURF$) by a factor ranging from 1.6 (first case uniform and directional illumination) to 3 (last case purely directional illumination). This clearly demonstrates the improved illumination invariance characteristics of the proposed method, especially for the difficult cases of non-uniform illumination.

In order to provide indirect comparison with other detectors that were not tested in our previous experiment, and at the same time have a reference point regarding the performance of the proposed algorithm, the widely known *Leuven* dataset was also used, consisting of several photographs of a parking lot captured under different illumination conditions [27]. Figure 12 depicts the respective graphs for this dataset. Similarly to the case of the *Phos* dataset, the proposed detector outperforms all the others, for the cases of repeatability (12(a)), overlap error (12(b)) and number of correspondences (12(c)).

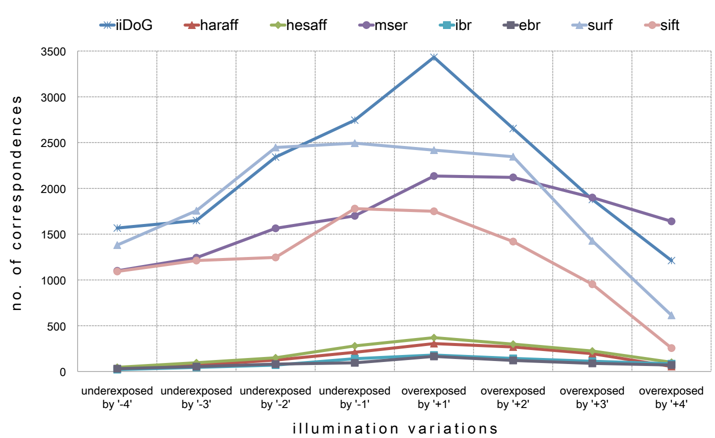
Since the main thrust of the proposed method is to locally equalize the gradient magnitude, in order to facilitate the thresholding of the extracted keypoints, one could argue that alter-ing the threshold of the classic *SIFT* detector could result to similar results. For this reason, we tested the detector performance of *iiDoG* and the classic *SIFT*, for various threshold values. The most challenging image (the one captured under only directional illumination lower right of Figure 9) was compared to the correctly exposed one (upper middle of Figure 9). After feature extraction by both detectors, a matching procedure took place where the amount of correct positive correspondences was measured. The feature extraction process was repeated for ten threshold values ranging from 0.01 to 0.3. The number of the detected key points of *iiDoG* and *SIFT*, during these threshold variations, is shown in Figure 13, while the number of correct positive matches is illustrated in Figure 14. The most interesting observation is the similar gradients of the lines both in key point detection and matching. Apparently, *iiDoG* demonstrates better performance than the original *SIFT* module for any threshold value. More importantly, for lower threshold values, the proposed detector exhibits double the number of correct matches, compared to *SIFT*. This increase in performance is a direct consequence of the fact that the proposed method detects keypoints also in the dark image regions, whereas SIFT does not. As a result, the number of correct matches, in



(a)



(b)



(c)

Figure 10. Evaluation of the proposed detector for various kinds of uniform illumination in *Phos* dataset; (a) Repeatability score for decreasing light; (b) Repeatability score for increasing overlap error; (c) Number of corresponding regions in the images.

the difficult case of non-uniform illumination with many underexposed image regions, is always higher for the *iiDoG* detector.

Figure 15 depicts the extracted matching points of a *DoG*-based *SIFT* algorithm and an *iiDoG*-based one, when applied to the same scene under uniform bright and non-uniform illumination. The results are depicted for different values of detector thresholds. In all cases, the proposed method exhibits greater number of matching points. Furthermore, the total number of matches remains more constant as the threshold value decreases. Finally, the *DoG*-based *SIFT* is more susceptible to wrong matches (lines which are not horizontal) compared to the proposed one.

6. Conclusions

This paper introduced a new operator combining the non-linear responses of center-surround cells of the *HVS*, as well as the reliability of the classic *DoG*. As a result, this new operator, *iiDoG*, exhibits increased output response in the underexposed image regions and the *DoG* response in any other cases. The operator can be used to create a scale-space, which in turn, can be a part of a *SIFT*-based detector module. The main advantage of this detector is the local equalization that the *iiDoG* operator introduces to the magnitude of gradient, according to which, contrast differences are boosted in the underexposed image regions, while kept intact in all other cases. Consequently, one global threshold can result in the extraction of keypoints, both in the dark and bright image regions.

Experimental results in different kinds and degrees of illumination demonstrated that the proposed approach outperforms existing detectors and exhibits constantly better results, especially in the difficult cases of uneven and non-uniform illumination. This kind of illumination conditions are quite usual in outdoor environments and can pose a considerable challenge to vision systems. Therefore, the increased illumination invariance of the proposed detector may be a solution to this problem. Additionally, the proposed method can be easily implemented, without requiring significant changes in the structure of existing *SIFT*-based systems. Finally, the fact that the output of the proposed detector is exactly the same with *DoG*, for the cases of well exposed image regions, ensures that the improvements introduced will only be targeted in shadows. Thus, no unpredictable or unwanted changes in performance will occur for the cases of correctly exposed images.

Acknowledgments

This study is partially supported by the research grant for ADSCs Human Sixth Sense Programme from Singapores Agency for Science, Technology and Research (A*STAR). Moreover, authors would like to express their gratitude to the reviewers of this paper for their valuable remarks.

References

[1] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 207(1167):187–217, 1980.

[2] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on*, 31(4):532–540, 1983.

[3] D.G. Lowe. Object recognition from local scale-invariant features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[4] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[5] X. Wang, L. Hou, and H. Yang. A feature-based image watermarking scheme robust to local geometrical distortions. *Journal of Optics A: Pure and Applied Optics*, 11(6):065401, 2009.

[6] S.A. Berrabah, H. Sahli, and Y. Baudoin. Visual-based simultaneous localization and mapping and global positioning system correction for geo-localization of a mobile robot. *Measurement Science and Technology*, 22(12):124003, 2011.

[7] Paolo Piccinini, Andrea Prati, and Rita Cucchiara. Real-time object detection and localization with sift-based clustering. *Image and Vision Computing*, 30(8):573 – 587, 2012.

[8] CREATE workshop. Official website of the create workshop. http://www.create.uwe.ac.uk/, June 2010.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[10] D. Westhoff, J. Zhang, and K. Huebner. Robust illumination-invariant features by quantitative bilateral symmetry detection. In *Information Acquisition, 2005 IEEE International Conference on*, page 6 pp., june-3 july 2005.

[11] Feng Tang, Suk Hwan Lim, N.L. Chang, and Hai Tao. A novel feature descriptor invariant to complex brightness changes. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2631 –2638, june 2009.

[12] Y. Yu, K. Huang, W. Chen, and T. Tan. A novel algorithm for view and illumination invariant image matching. *Image Processing, IEEE Transactions on*, 21(1):229–240, 2012.

[13] S. Lee, S. Lee, and J.J. Yoon. Illumination-invariant localization based on upward looking scenes for low-cost indoor robots. *Advanced Robotics*, 26(13):1443–1469, 2012.

[14] Anindya Roy and Se?bastien Marcel. Haar local binary pattern feature for fast illumination invariant face detection. In *Proc. British Machine Vision Conference*, pages 19.1–19.12, 2009.

[15] X. Cao, W. Shen, L. G. Yu, Y. L. Wang, J. Y. Yang, and Z. W. Zhang. Illumination invariant extraction for face recognition using neighboring wavelet coefficients. *Pattern Recognition*, 45(4):1299–1305, April 2012.

[16] Paul R. Martin and Ulrike Grunert. Gaglion cells in mammalian retinae. In J S Werner and L M Chalupa, editors, *The Visual Neurosciences*, volume 1 of *Bradford Books*, chapter 26, pages 410–421. Mit Press, 2004.

[17] S. Grossberg. Visual boundaries and surfaces. In J S Werner and L M Chalupa, editors, *The Visual Neurosciences*, volume 2 of *Bradford Books*, chapter 26, pages 1624–1639. Mit Press, 2004.

[18] S. Grossberg and D. Todorovic. Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena. *Attention, Perception, & Psychophysics*, 43(3):241–277, 1988.

[19] L. Pessoa, E. Mingolla, and H. Neumann. A contrast-and luminance-driven multiscale network model of brightness perception. *Vision Research*, 35(15):2201–2223, 1995.

[20] K.I. Naka and W.A.H. Rushton. S-potentials from luminosity units in the retina of fish (cyprinidae). *The Journal of physiology*, 185(3):587–599, 1966.

[21] T. Duong and R.D. Freeman. Contrast sensitivity is enhanced by expansive nonlinear processing in the lateral geniculate nucleus. *Journal of neurophysiology*, 99(1):367–372, 2008.

[22] GRJ Cooper. Balancing images of potential-field data. *Geophysics*, 74(3):L17–L20, 2009.

[23] Chris Wijns, Carlos Perez, and Peter Kowalczyk. Theta map: Edge detection in magnetic data. *Geophysics*, 70(4):L39–L43, 2005.

[24] J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1):103–112, 2005.

[25] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005.

[26] Phos database. Official website of the proposed database. http://utopia.duth.gr/~dchrisos/pubs/database2.html, June 2012.

[27] Leuven dataset. Official website of the leuven dataset. http://www.robots.ox.ac.uk/vgg/research/affine/index.html, June 2012.

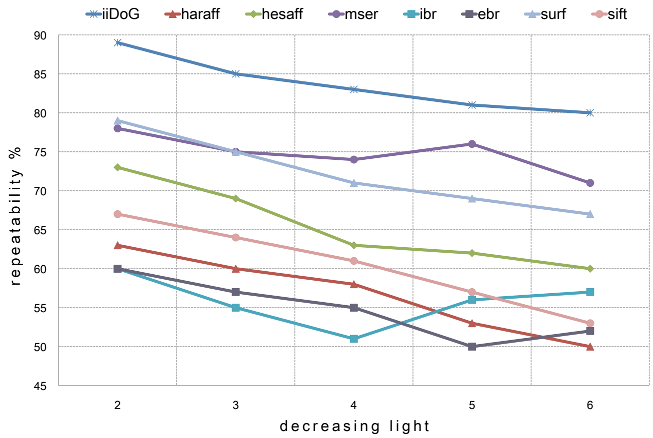
[28] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(3):177–280, 2008.

[29] T Tuytelaars and L van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *Proc. British Machine Vision Conference*, volume 2, pages 38.1–38.14, 2000.

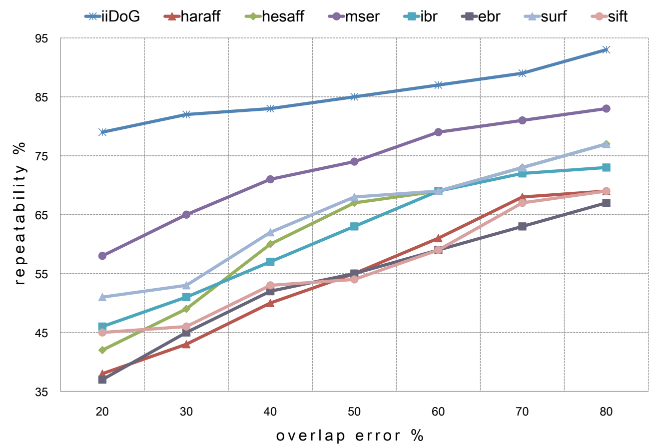
[30] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[31] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Computer Vision, ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 128–142. Springer Berlin Heidelberg, 2002.

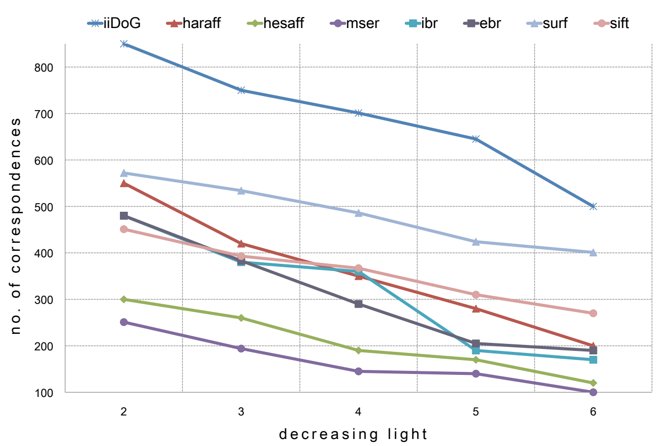
[32] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International journal of computer vision*, 59(1):61–85, 2004.



(a)

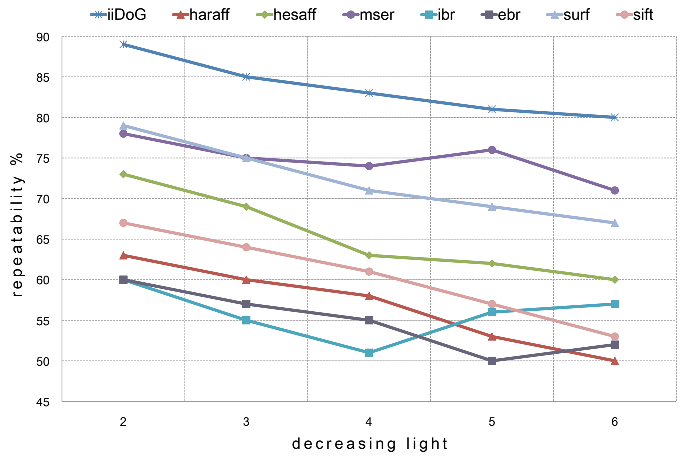


(b)

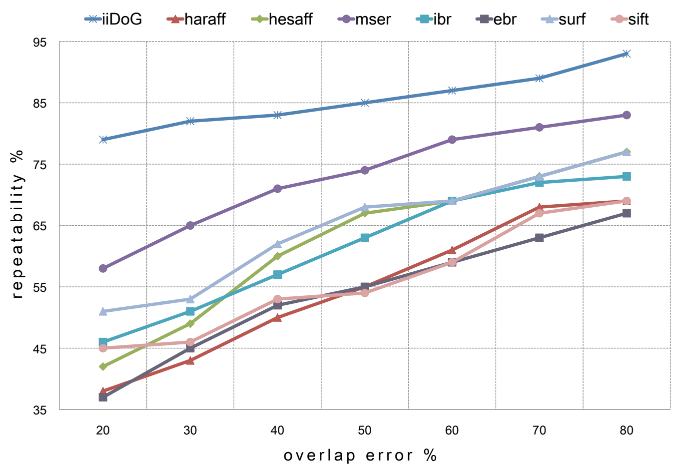


(c)

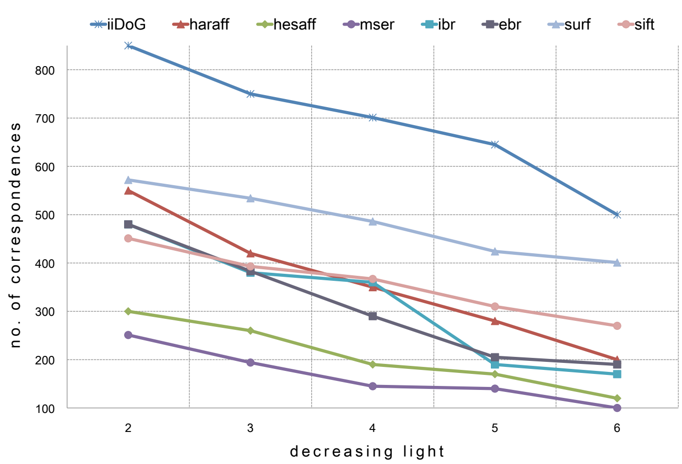
Figure 11. Evaluation of the proposed detector for various degrees of non-uniform illumination in *Phos* dataset; (a) Repeatability score for decreasing light; (b) Repeatability score for increasing overlap error; (c) Number of corresponding regions in the images.



(a)



(b)



(c)

Figure 12. Evaluation of the proposed detector for the Leuven sequence; (a) Repeatability score for decreasing; (b) Repeatability score for increasing overlap error; (c) Number of corresponding regions in the images.

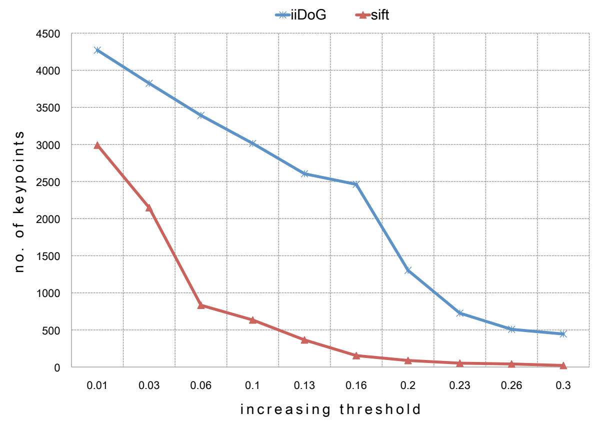


Figure 13. Number of detected keypoints between *iiDoG* and the detector module of *SIFT* for various threshold values.

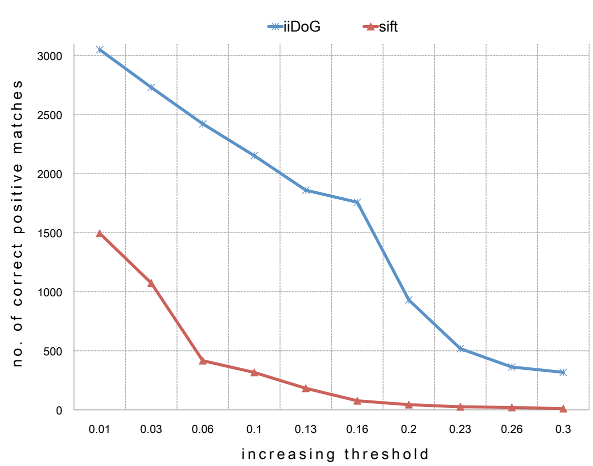


Figure 14. Number of correct positive matches between *iiDoG* and the detector module of *SIFT*, for various threshold values.

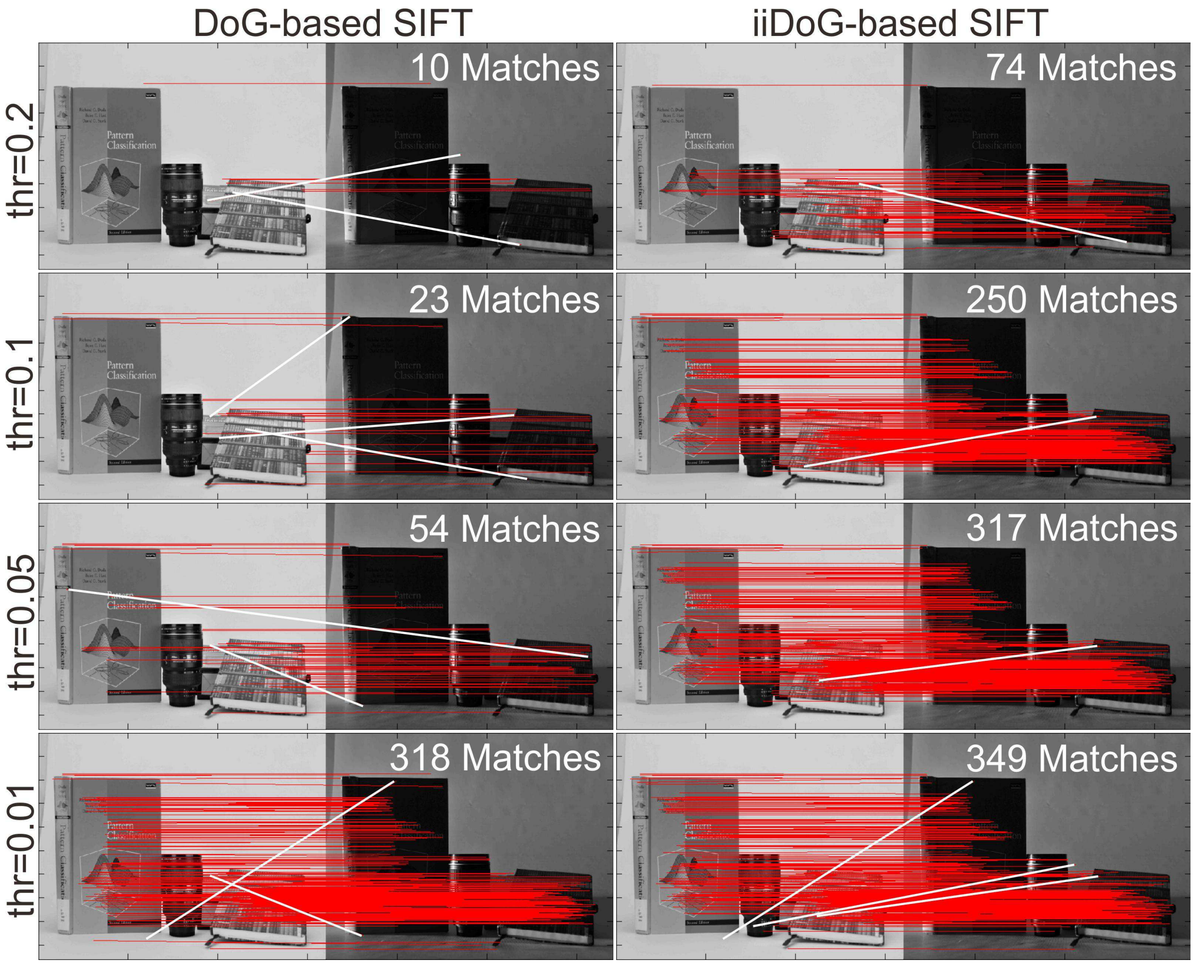


Figure 15. Comparison of matching points between a *DoG*-based *SIFT* and the proposed *ii-DoG*-based *SIFT*.