Improving the Robustness in Feature Detection by Local Contrast Enhancement

Vassilios Vonikakis
Advanced Digital Sciences Center
Singapore 138632
Email: bbonik@adsc.com.sg

Dimitris Chrysostomou Rigas Kouskouridas and Antonios Gasteratos
Democritus University of Thrace
Xanthi Greece 67100
Emails: dimitris.chrysostomou@ieee.org rkouskou@pme.duth.gr gasteratos@ieee.org

***Abstract*—This paper presents a new feature detector, with improved local contrast performance. The proposed method is based on an improved non-linear version of the classic Difference of Gaussians, which exhibits increased sensitivity to low contrast. Additionally, it does not employ computationally expensive or memory demanding routines. In order to evaluate the degree of illumination invariance that the proposed, as well as, other existing detectors exhibit, a new benchmark image database has been created. 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 existing ones. These results are evident for both uniform and non-uniform illumination, evincing a favorable usage of the proposed feature detector by robotic platforms working in outdoor working environments.**

I. INTRODUCTION

Contemporary robotics systems are expected to provide applicable services in humans' daily life, which can only be achieved if they are equipped with advanced vision systems that can cope with any lighting condition. For the adequate accomplishment of several challenging tasks, such as: object manipulation, obstacle avoidance and navigation, 3D reconstruction and action recognition, visual feedback is a prerequisite. Towards this end, several methods that incorporate computer vision algorithms in robotics applications were recently presented [1], [2], [3], [4]. However, in real life situations optimal illumination circumstances seldom exist. Thus, robotic platforms operating in such working environments may suffer from low quality input information. Highly motivated from that fact, we propose a non-linear Difference of Gaussians (DoG) detector that aims to provide feature extraction results that are closer to optimum illumination settings. DoG is one of the most widely used operators in many vision domains. In biological inspired vision, the DoG operator is used to model the receptive fields of ganglion cells of the retina [5]. In computer vision, it is used either for edge extraction, since it approximates the Laplacian of Gaussian (LoG), or for feature detection, by applying it on the different scales of a Gaussian pyramid. The latter is the basis of the SIFT detector [6], which is extensively used in object and scene recognition applications.

Although widespread, the DoG operator is significantly affected by illumination and low local contrast. Fig. 1b depicts the results of the DoG operator on an image of a textured surface under a strong shadow (Fig. 1a). It is evident that the output of the operator is higher at the well illuminated regions of the image, whereas it decreases dramatically at the shadowed ones. Similar are the results in Fig. 1d, where a scene with foggy background is depicted; the output of the DoG operator is considerably higher at the well-defined foreground, than at the hazy background (Fig. 1e). Since the DoG operator is affected by the gradient of the image, low local contrast conditions (e.g. foggy/hazy scenes or shadows/highlights) result to low DoG outputs. This can negatively affect the performance of systems, which use DoG-based feature detectors, especially when operated outdoors, where the imaging conditions cannot be controlled.

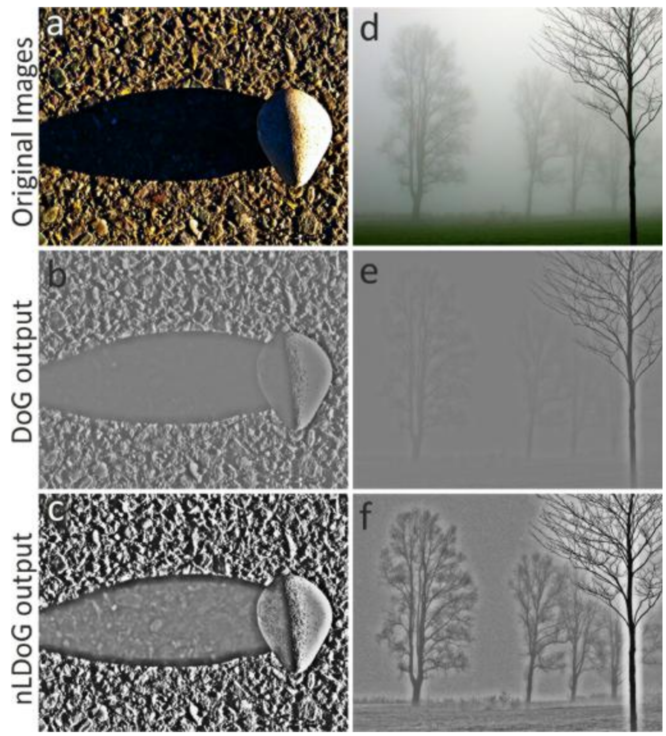

Fig. 1. Results of the classic DoG operator and of the proposed method for scenes with challenging illumination conditions.

During the past decade, remarkable efforts were made to build new vision frameworks for robust object recognition

978-1-4577-1775-8/12/$26.00 ©2012 IEEE

in cluttered environments. The techniques developed so far are mostly based on appearance features with spatial estate. Algorithms of this field extract features with local extent that are invariant to possible illumination, viewpoint, rotation and scale changes [6], [7]. The two main sub-mechanisms of such frameworks are the detectors and descriptors of areas of interest, respectively. The main idea behind interest point detectors is the pursuit of points or regions in a scene containing data that are salient within their local neighborhood. In a following step, a 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.

The DoG-based feature detector was initially introduced in [8], where interest point extraction is accomplished by the LoG and several other derivative based operators. A more effective solution to the problem of local 3D extrema estimation has been presented in [9]. Generally, DoG-based detectors are reported to be invariant to scale and affine changes opposite to illumination alternations [10], [11]. In this paper we study comparatively the efficiency of the proposed non-linear DoG with several state of the art detectors namely (a) the 'Maximally Stable Extremal Region' detector $(mser)$ [12]; (b) the 'Harris-Affine' detector [10], [13], [14]; (c) the 'Hessian-Affine' detector [10], [13]; (d) the 'intensity extrema-based region detector' (ibr) [15] and (e) the 'edge-based region detector' (ebr) [16]. The aforementioned detectors are compared on several datasets that are available in [17], [18], [19].

The main objective of the proposed algorithm is to increase the performance of the classic DoG detectors, thus, to provide detection robustness in low local contrast occasions. The classic Gaussian pyramid is still used for the decomposition of the image into different scales. However, the simple difference between adjacent scales, i.e. the classic DoG operator, is substituted by a non-linear function, which exhibits increased response to the small local differences between adjacent scales. Furthermore, the proposed approach does not require computationally expensive or memory demanding techniques to be implemented. As a result, the proposed function significantly increases the amount of extracted features in scenes, compared to the classic DoG detector. This increase in performance is considerably higher in images with low local contrast, making the proposed detector particularly appropriate for use in outdoor imaging applications, where lighting conditions cannot be easily controlled. In order to evaluate the performance of the proposed method, a new benchmark image database is created, containing scenes with real objects, under various degrees and combinations of uniform and non-uniform illumination. As a result, the new dataset allows the extensive testing of existing detector/descriptor algorithms, evaluating specifically their performance under various illumination conditions.

The remainder of the paper is organized as follows: Section II describes the proposed novel non-linear DoG detector. Section III presents the new benchmark database. The experimental results are presented in Section IV and concluding remarks are made in Section V.

II. Proposed DoG Feature Detector

The classic DoG feature detector, as described in [6], is essentially a subtraction between adjacent scales of a Gaussian pyramid of an image, as (1) indicates:

$$\begin{aligned} DoG(i,j,\sigma) &= \big[G(i,j,\kappa\sigma) - G(i,j,\sigma)\big] * I(i,j) \\ &= L(i,j,\kappa\sigma) - L(i,j,\sigma) \end{aligned} \tag{1}$$

where I is the input image, DoG is the difference of Gaussians for standard deviation σ, G is the Gaussian function, L is the blurred image resulted by the convolution of I and G, (i,j) are the spatial coordinates and κ is a multiplicative factor, that determines the different levels of blurring between adjacent scales. $DoG(i,j,\sigma)$ is a local contrast indicator at position (i,j) and for standard deviation σ. Yet as discussed in Section I, classic DoG is significantly affected by lighting conditions, resulting frequently into low local responses, due to illumination.

In order to overcome similar problems, a biologically-inspired contrast enhancement method was presented for intensity values [20]. It employed two mapping functions for increasing the response to local contrast, extracted by center-surround receptive fields. Since DoG is essentially a center-surround operator, this approach can be also adopted in DoG-based feature detectors. Therefore, in order to increase the DoG response to low local contrast stimuli, the following mapping function is employed [20], [21]:

$$f(x) = \frac{x(A+B)}{x+A}, x \in [0,B], A \in (0,+\infty) \tag{2}$$

where B is the maximum value of x and A is a factor determining the nonlinearity degree of the mapping. Equation (2) maps x from the interval $[0,B]$ to $[0,B]$ in a nonlinear fashion determined by A (Fig. 2). Small A values, relatively to B, result to a steep nonlinear mapping. Furthermore, as A approaches zero, $f(x)$ tends to asymptotically resemble a right angle. The higher the value of A, the more $f(x)$ asymptotically resembles linearity. Practically, values of A, five times greater than B, result to an almost linear mapping.

For an image I with possible pixel values in the interval $[0,B]$ (usually B=255) all DoG values will be within the interval $[-B,B]$. Consequently, the absolute value $|DoG|$ will be within the range $[0,B]$. As a result, it is possible to substitute x with $|DoG|$ in (2), and obtain:

$$f(|DoG(i,j,\sigma)|) = \frac{|DoG(i,j,\sigma)|(A+B)}{|DoG(i,j,\sigma)|+A} \tag{3}$$

Equation (3) maps the absolute value of the local contrast in a non-linear fashion, as extracted by the DoG operator, to the interval $[0,B]$, increasing the low local contrast values according to the nonlinearity factor A. The absolute value of the local contrast however, is of little use to a feature detector, since it lacks useful information about the sign of the local contrast. Consequently, it is very important to

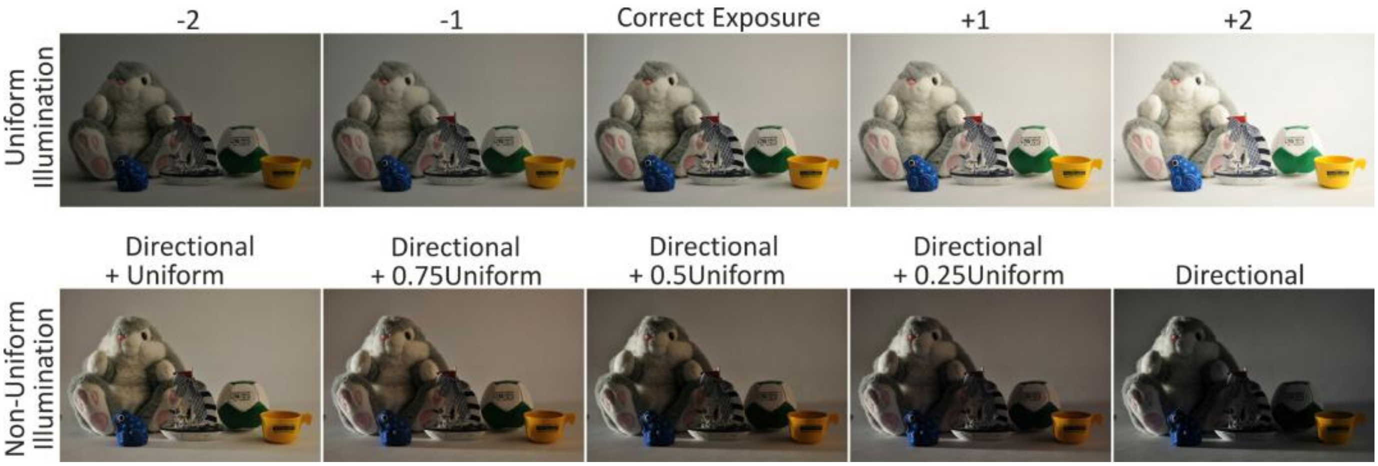

Fig. 3. One of the scenes of the new benchmark image database.

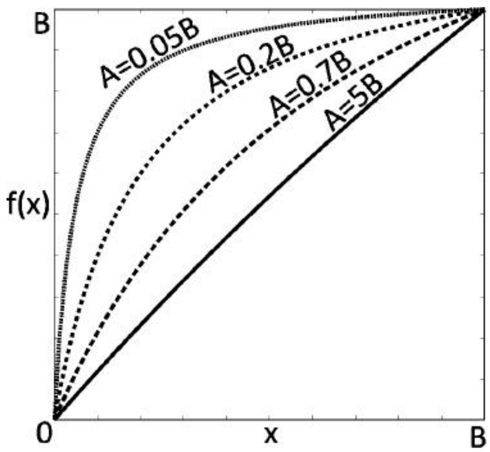

Fig. 2. The mapping function employed for increasing the response to low local contrast of the DoG detector.

retain the sign of the original DoG operator. In order to do so, (3) is transformed into (4), which performs the same nonlinear mapping, increasing the low local contrast values, while preserving the sign of the DoG operator.

$$nLDoG(i,j,\sigma) = \frac{DoG(i,j,\sigma)(A+B)}{|DoG(i,j,\sigma)|+A} \tag{4}$$

where $nLDoG$ is the proposed nonLinear DoG operator. Next, the proposed feature detector applies (4) into the scales of a Gaussian pyramid, extracting features at multiple scales. The use of the nonlinear DoG operator results to a greater number of detected features, due to the increased sensitivity of the operator to low local contrast.

Moreover, equation 4 does not increase significantly the computational burden of the detector, since DoG, which is still the main factor in the proposed $nLDoG$, has to be also computed in the classic DoG detector. Apart from this, the proposed $nLDoG$, requires two extra additions, one multiplication and one division. Taking into account that B is a constant and A can be set to a specific value, equation 4 can be precomputed with a Look-up-table (LUT) minimizing the computational cost of the proposed $nLDoG$. Fig 1c and Fig. 1f depict the results of the proposed $nLDoG$ detector when applied to the original images of Fig. 1a and Fig. 1d respectively, in a single spatial scale, and for $A = 0.1B$. It is apparent that the proposed detector extracts more details compared to the classic DoG, when applied in the same spatial scale.

III. *Phos* Benchmark Image Database

In order to test the proposed feature detector, 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 [18], [19], 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. All the previously presented tests, regarding illumination invariance, are done by manually adjusting image brightness with image processing software.

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. Moreover, 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, which is the case for many robotic systems, 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 the gap of the existing benchmark databases, by specializing to realistic illumination conditions. More particularly, every scene of the database contains 10 different images: 5 images captured under various strengths of uniform illumination, and 5 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. Single objects scenes were also captured under the same illumination variations. Fig. 3 depicts one scene from the new image database. $Phos$ database is available to the public at [17].

Uniform illumination (first row of Fig. 3) 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 -2 and +2 stops from the original correctly exposed image. Thus, for every scene there are two underexposed and two overexposed images with uniform illumination.

Non-uniform illumination (second row of Fig. 3) 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, five 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 deteriorates the performance of the detectors.

IV. 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 [17] and in the $Leuven$ sequence provided in [11].

A. 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. The evaluation procedure followed is similar to [11], 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 so as the ground truth measurement of the possible transformation to be calculated. 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 $(1 - interesection/union)$ 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.

B. Test data and results

In order to demonstrate the effect of the non-linearity factor A into the feature extraction procedure, the proposed detector was tested using three different values of A, relatively to B. $nLDoG - 1$ stands for the proposed detector with A value equals to $0.01B$, $nLDoG - 2$ stands for $A = 0.2B$ and $nLDoG - 3$ stands for the case where the proposed detector tested with $A = 0.4B$. The value of 0.01B was selected, since it represents a very steep non-linear mapping, which significantly increases low contrast values. Smaller A values than 0.01B, would result to an even steeper non-linear mapping. This however, would not necessarily result to a considerable increase in the number of detected features, since, as A approaches 0, the mapping curve asymptotically resembles a right angle. In other words, for very small A values, there are even smaller differences in the mapping curve, which will inevitably result to similar feature detection numbers. The value of 0.4B was selected since it is indicative of a threshold, above which, the mapping function acquires a more gentle slope, resulting to decreased enhancement of low contrast details, as indicated by our experiments. Finally, value 0.2B was selected as an intermediate step between 0.01B and 0.4B, in order to demonstrate the gradual changes in performance that the parameter A introduces. Additionally, different detectors were tested for their robustness and repeatability scores: the 'Maximally Stable Extremal Region' - $(mser)$ detector [12], the 'Harris-Affine' - $(haraff)$ detector [13], the 'Hessian-Affine' - $(hesaff)$ detector [13], the 'intensity extrema-based region detector' - (ibr) [15] and the 'edge-based region detector' - (ebr) [16].

Figure 4 presents the repeatability score for the case of uniform illumination in the $Phos$ dataset, depicted in the first sequence of Fig. 3. The image retaining the correct exposure conditions was used as reference and each one of the others (+2, +1, -1, -2) as subjects for comparison.

In the ideal case the repeatability line is horizontal. These ideal illumination conditions, however, exist only in artificial scenes. Fig. 4(a) depicts the relative number of corresponding regions for an overlap error of 40%. As it is expected, the performance of all the evaluated feature detectors is not ideal. However, the proposed modification of the DoG detector clearly outperforms all the others. Figure 4(b) displays the actual number of corresponding regions with overlap error of 40%.

The proposed method obtains the best repeatability score as well as the highest number of corresponding regions in most examined cases. The improvement of the robustness in feature detection is apparent as the proposed method exhibits almost optimal results along any illumination variations. Figure 4(c) shows the repeatability score computed for the most difficult pair of images, i.e. the reference one and the one with the most challenging light alterations. The repeatability score for all the detectors is tested as the overlap error increases. Also in this case the proposed detector clearly exhibits the highest repeatability scores, under any lighting condition.

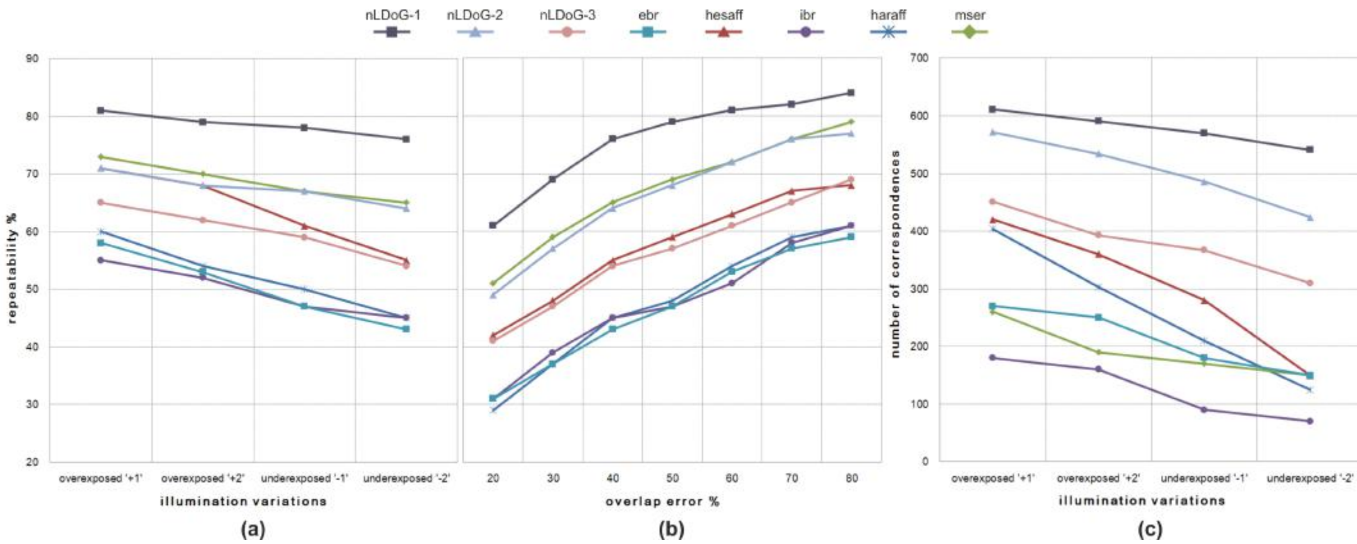

Fig. 4. Evaluation of the proposed detector. (a) Repeatability score for decreasing light in $Phos$ dataset with uniform illumination. (b) Repeatability score for increasing overlap error. (c) Number of corresponding regions in the images.

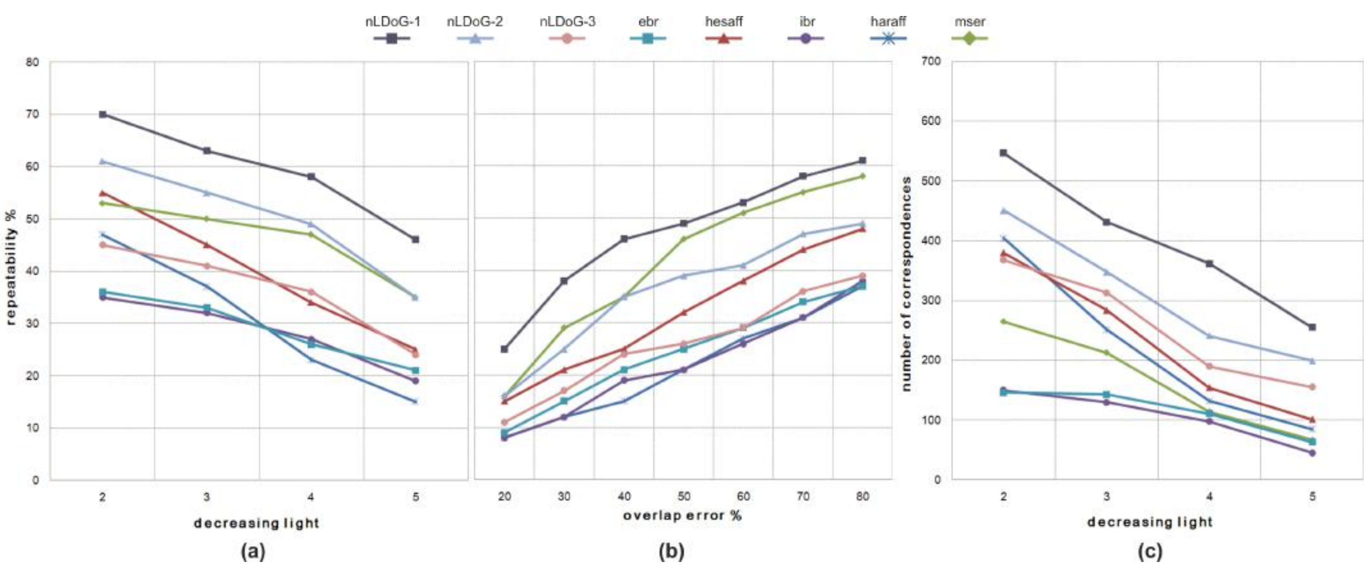

Fig. 5. Evaluation of the proposed detector. (a) Repeatability score for decreasing light in $Phos$ dataset with non - uniform and directional illumination. (b) Repeatability score for increasing overlap error. (c) Number of corresponding regions in the images.

Similar performance is achieved at the other two examined sequences, i.e. the $Phos$ dataset with non-uniform illumination, (Fig 5), depicted in the second row of Fig 3, and the $Leuven$ sequence (Fig 6). In both of these cases, the proposed detector, and particularly $nLDoG - 1$, outperforms all the existing detectors, exhibiting high repeatability scores with many corresponding regions at the same time.

From the experiments in Figures 4, 5, and 6 some important conclusions can be derived. First, it is evident that lower A values, relatively to B, increase the performance of the detector. The experimental tests showed that a value of $A = 0.01B$ exhibits improved results for all kinds of scenes and illumination conditions. Second, for uniform illumination the chart morphology is independent to the dataset i.e. the charts are similar either applied to $Leuven$ (Fig. 6) or $Phos$ dataset. This proves that the technique of adjusting the image brightness, adopted until now, accounts only for uniform illumination changes. On the contrary, the non-uniform illumination dataset of the $Phos$ database (Fig. 5) has a different impact on the performance of all detectors, decreasing their correspondences more rapidly, as local illumination becomes dimmer. Therefore the $Phos$ dataset can be used for evaluating detector performance in more realistic lighting conditions.

V. Conclusions

In this paper a modified, non linear version of the DoG detector has been presented. An experimental evaluation of the proposed detector with other popular detectors has been conducted for sequences of both uniform and non-uniform

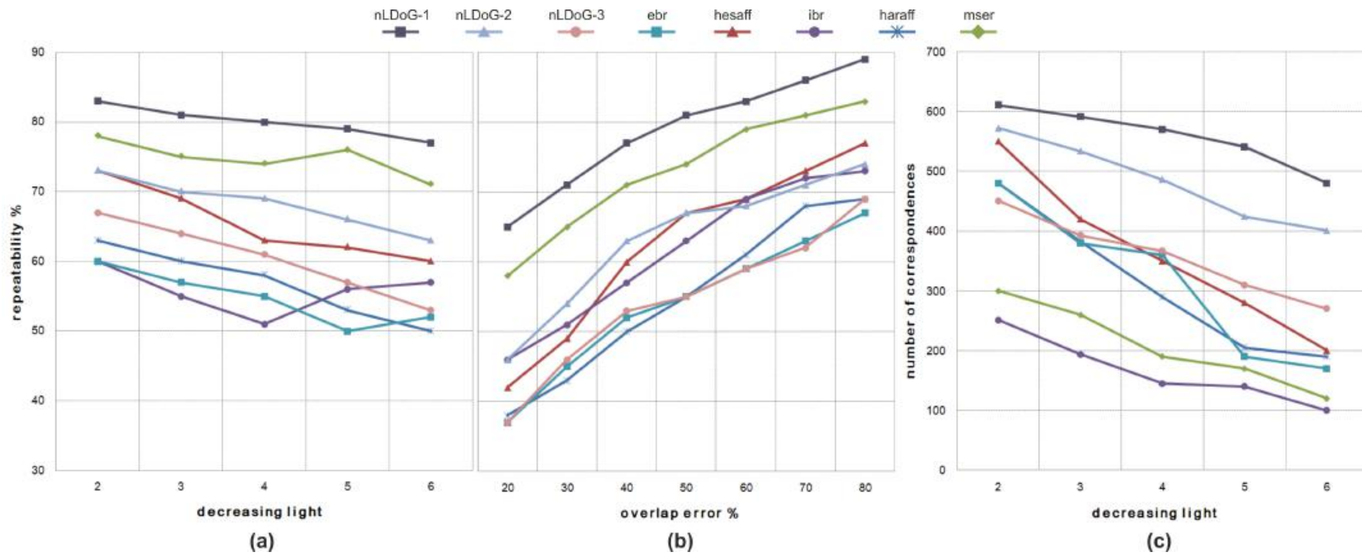

Fig. 6. Evaluation of the proposed detector. (a) Repeatability score for decreasing light in $Leuven$ sequence. (b) Repeatability score for increasing overlap error. (c) Number of corresponding regions in the images.

illumination conditions. In all tests the highest repeatability score was obtained by $nLDoG-1$ followed closely by $mser$. The comparison has shown that the performance of all presented detectors declines slowly, with similar rates, as the light in the images decreases. The $nLDoG-1$ clearly outperforms the other detectors both in terms of repeatability score and number of corresponding regions in sequences with either uniform or non - uniform illumination. In practice, this increase in performance, means that the proposed detector will extract greater number of correct features in non-ideal lighting conditions.

Moreover, the proposed detector is computationally inexpensive and requires less memory resources, as it saves an extra preprocessing stage prior to the application of the detector. Consequently, the proposed approach represents a suitable choice for mobile robotics applications and especially for outdoor or space scenarios, where severe illumination changes may occur.

ACKNOWLEDGMENT

Authors would like to thank Dr. George Pavlidis and the staff in Clepsydra center for their help and support during the construction of the $Phos$ image database.

REFERENCES

[1] H. Kjellström, J. Romero, and D. Kragic, "Visual object-action recognition: Inferring object affordances from human demonstration," *Computer Vision and Image Understanding*, 2010.

[2] B. Rasolzadeh, M. Björkman, K. Huebner, and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in the real world," *The International Journal of Robotics Research*, vol. 29, no. 2-3, p. 133, 2010.

[3] R. Kouskouridas, A. Amanatiadis, and A. Gasteratos, "Guiding a Robotic Gripper by Visual Feedback for Object Manipulation Tasks," in *IEEE International Conference on Mechatronics*. IEEE, 2011.

[4] L. Nalpantidis and A. Gasteratos, "Stereo vision for robotic applications in the presence of non-ideal lighting conditions," *Image and Vision Computing*, vol. 28, no. 6, pp. 940–951, 2010.

[5] P. Martin and U. Grünert, "Ganglion cells in mammalian retinae," *The visual neurosciences*, pp. 410–421, 2004.

[6] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[8] T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79–116, 1998.

[9] D. Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision*. Published by the IEEE Computer Society, 1999, pp. 1150–1157.

[10] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International journal of computer vision*, vol. 60, no. 1, pp. 63–86, 2004.

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1, pp. 43–72, 2005.

[12] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[13] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proceedings of the 7th European Conference on Computer Vision-Part I*. Springer-Verlag, 2002, pp. 128–142.

[14] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or How do I organize my holiday snaps?," *Computer VisionECCV 2002*, pp. 414–431, 2002.

[15] T. Tuytelaars and L. Van Gool, "Wide baseline stereo matching based on local, affinely invariant regions," in *British Machine Vision Conference*. Citeseer, 2000, pp. 412–425.

[16] T. Tuytelaars and L. Van Gol, "Matching widely separated views based on affine invariant regions," *International journal of computer vision*, vol. 59, no. 1, pp. 61–85, 2004.

[17] "http://utopia.duth.gr/ dchrisos/pubs/database.html/."

[18] "http://www.robots.ox.ac.uk/ vgg/research/affine/index.html."

[19] J. Geusebroek, G. Burghouts, and A. Smeulders, "The Amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.

[20] V. Vonikakis and I. Andreadis, "Multi-scale Image Contrast Enhancement," in *10th International Conference on Control, Automation, Robotics and Vision*. IEEE, 2008, pp. 856–861.

[21] V. Vonikakis and I. Andreads, "Fast automatic compensation of under/over-exposured image regions," *IEEE Pacific Rim Symposium on Image and Video Technology*, pp. 510–521, 2007.