Luminance Ratios for Stable SIFT Keypoint Detection in Challenging Illuminations Scenarios

Lucas Niquele Endrice and Hugo Vieira Neto
Graduate Program in Electrical and Computer Engineering
Federal University of Technology - Paraná
Curitiba, Paraná
Email: lucasendrice@alunos.utfpr.edu.br, hvieir@utfpr.edu.br

***Abstract*—Invariance to changes in scene illumination is a desirable feature in computer vision, but its accomplishment is still a challenge even for modern systems. This paper proposes a novel way to detect keypoints in scale-space by exploring the concept of luminance ratio first introduced in Wallach's Lightness Constancy Theory, instead of using linear difference operators. The hypothesis is that an operator based on the concept of luminance ratio may provide more robust results with respect to illumination changes. In order to test this hypothesis, the SIFT keypoint detector was adapted to the proposed approach with minimum implementation effort. Experimental results show that the luminance ratio approach indeed yields more stable keypoint detection than the original SIFT based on linear difference operators, specially in lower illumination conditions. Existing SIFT implementations can be easily adapted to use the luminance ratio concept for improved operation in environments with uncontrolled lighting conditions.**

I. Introduction

Feature extraction from images obtained from real world environments is a challenging task even for modern computer vision systems. Adaptation to different illumination conditions is natural for the human visual system, but may be a problem for artificial systems. Historically, edges in images have been estimated by operators such as the Difference of Gaussians (DoG). Despite being computationally fast, difference operators may impact directly in the final performance of the artificial visual system when illumination conditions are variable [1], [2].

A solution for this quest for robustness regarding illumination may rely on the concept of *lightness constancy*, illustrated in Figure 1. In this example, the indoor illumination intensity is ten times weaker than the outdoor intensity, but the human visual system is able to perceive both conditions as being virtually the same. If contrast perception is based on luminance differences, it results in 80 intensity units indoors and 8000 units outdoors, which constitutes a huge variation in artificial "perception". But if luminance ratios are considered, one can notice that the result is the same (9:1) for both scenarios, which seems a reasonable path to be followed [3, p. 125].

In order to evaluate the luminance ratio hypothesis, which was presented by Hans Wallach in [4], by comparing ratio-based and difference-based performances in changing illumination scenarios, the SIFT keypoint detector [5], [6] was used. The expectation was that the ratio-based approach would result in more stable keypoint localization with respect to changes in illumination, especially in darker conditions, since SIFT already presents some level of illumination invariance.

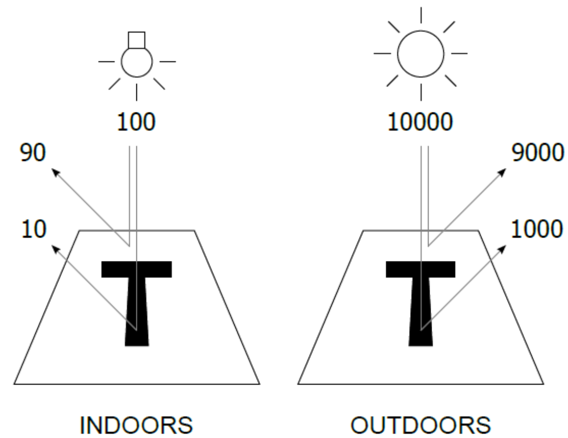


Fig. 1. Lightness constancy concept. Human perception of contrast does not change despite different illumination intensities (adapted from [3]).

A brief literature review of the SIFT keypoint detection framework is given in II, while an adaptation to incorporate the ratio-based approach is described in Section III. The experimental setup is of the work presented in Section IV, while the results obtained are presented and discussed in Section V. In Section VI conclusions are drawn and future work outlined.

II. Literature Review

Several computer vision methods rely on luminance differences from gray level images to extract information about their content, such as the locations of edges and corners. However, this approach may be affected by illumination changes.

Hans Wallach presented a possible solution to the illumination invariance problem with his Lightness Constancy Theory (or brightness constancy, as he called it) [4]. Lightness constancy is the perception of a given achromatic surface having the same lightness despite of differences in illumination or viewing conditions [3, p. 125].

According to Wallach, contrast perception is determined by the ratio between the brightness of two regions. He tested his theory conducting the experiment illustrated in Figure 2, in which there were two separate white screens and two light

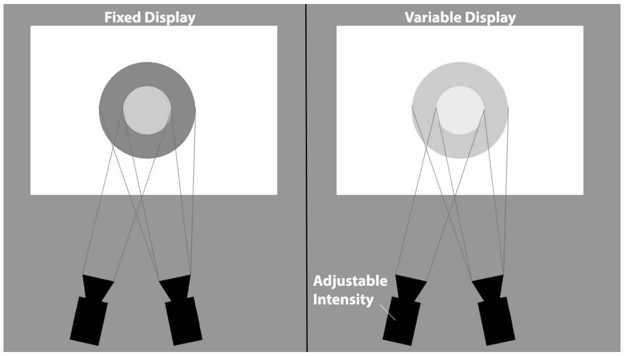


Fig. 2. Wallach's experiment on contrast perception. The goal was to match the adjustable center intensity to the fixed one and get insight on how lightness contrast is perceived (adapted from [3]).

sources for each screen, one of them projecting a circular region and the other projecting a surrounding ring [4].

The sources in the fixed screen were set in different fixed intensities throughout the experiment, and the ring source in the variable screen was set differently from the ring source in the fixed screen. The resulting fixed and variable center/ring displays were shown to human subjects who were told to adjust the intensity of the variable center source to match the contrast perception of the fixed center source [4].

Based on the results, Wallach concluded that contrast perception depends in close approximation on the ratios of the intensities involved and seems to be independent of the absolute intensity of local stimulation [4]. The results achieved by Wallach were later corroborated by Jacobsen and Gilchrist [7], and currently support the ratio-based contrast theory of lightness constancy for the human visual system.

Vieira Neto [2] exploited Wallach's ratio-based theory using the Sobel edge detection in preliminary experiments involving synthetic changes in illumination (darkening only). The results obtained showed that the ratio-based approach was more stable regarding illumination changes than the original difference-based Sobel edge detector. In the present work, a more sophisticated scenario is explored by experimenting the ratio-based approach with the SIFT keypoint detector in real world illumination conditions.

Despite a vast literature regarding improvements in SIFT, most efforts have focused on keypoint description, while the keypoint detection framework is almost kept untouched. Perhaps the most important study involving SIFT keypoint detection under illumination changes was conducted by Vonikakis and colleagues [1], in which an illumination invariant operator in scale-space was experimented in the Phos dataset, which was specially conceived for this purpose.

The operator proposed in [1] was inspired by the nonlinear responses of center-surround cells of the human visual system and by the reliability of the classic Difference-of-Gaussians (DoG). The approach combines DoG and *n*DoG operators into one piecewise function – the original DoG is used for higher contrast regions, while the *n*DoG operator for lower contrast regions. Equation 1 represents the piecewise *ii*DoG function:

$$iiDoG = \begin{cases} nDoG = \frac{S-C}{S+C} & if\ C+S < B \\ 0 & if\ C = S = 0 \\ DoG = \frac{S-C}{B} & if\ C+S > B, \end{cases} \quad (1)$$

in which S is the surround, C the center and B is the maximum value that S or C may take.The results obtained with the *ii*DoG operator show better resilience to illumination changes than the original SIFT and other detectors. Better repeatability scores and number of correspondences were achieved in most experiments.

Another study conducted by Boonsivanon and Meesomboon [8] proposes the use of morphological filtering to improve SIFT keypoint detection in different illumination scenarios, using the Phos dataset once more. However, that study turns out to be rather inconclusive because it focuses only in the absolute number of detected keypoints regardless of their repeatability throughout illumination changes.

Huang and colleagues [9] have used Laplacian-of-Bilateral (LoB) filtering, instead of the conventional DoG, for the construction of the SIFT scale-space. It is argued in that study that LoB yields more repeatable keypoints with respect to changes in illumination conditions, but in a very limited number of images from the Mikolajczyk dataset [10] concerning varying illumination.

III. Proposed Ratio-based Approach

The original SIFT approach [5], [6] detects stable keypoints by subtracting two nearby levels in a scale-space, constructed by Gaussian filters with increasing widths (by a factor k) convolved with the input image, as shown in Equation 2:

$$\begin{aligned} D(x,y,\sigma) &= G(x,y,k\sigma) * I(x,y) - G(x,y,\sigma) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma), \end{aligned} \quad (2)$$

in which $I(x,y)$ is the input image, $G(x,y,\sigma)$ is the Gaussian filter kernel with σ width, $L(x,y,\sigma)$ is the Gaussian-filtered version of the image and $D(x,y,\sigma)$ is the DoG response obtained from the subtraction of two nearby levels in scale-space.

A straightforward way to obtain ratio-based keypoint detection within the SIFT framework is by using the property of logarithms expressed in Equation 3:

$$log\left(\frac{x}{y}\right) = log(x) - log(y). \quad (3)$$

Therefore, it is possible to use Eqs. (2) and (3) in order to implement the ratio-based keypoint detection with minimum changes in the original SIFT. And so, the proposed approach is based on Equation 4:

$$L_l(x,y,\sigma) = \frac{log[(N-1)[G(x,y,\sigma) * I(x,y)] + 1]}{log(N)}, \quad (4)$$

in which $L_l(x,y,\sigma)$ is the logarithmic mapping from the scale-space levels to the range [0, 1] and N is the base of the

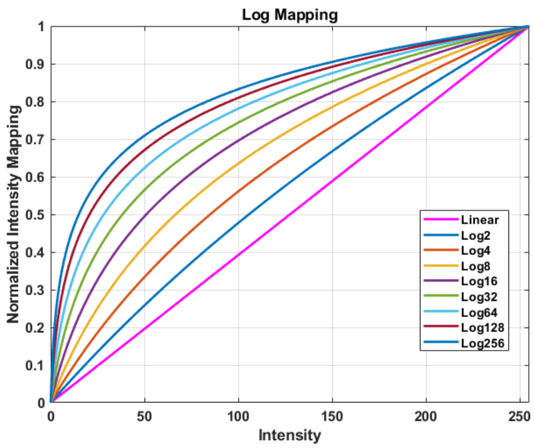


Fig. 3. Logarithmic mapping for the different values of N.

logarithm to be used. Figure 3 shows the resulting 8-bit intensity value mapping for $N = \{2, 4, 8, 16, 32, 64, 128, 256\}$ – as can be noticed, the higher the value of N, the higher the nonlinearity of the mapping. The logarithmic Ratio-of-Gaussians function is finally given by Equation 5:

$$R_l(x, y, \sigma) = L_l(x, y, k\sigma) - L_l(x, y, \sigma). \quad (5)$$

The implementation consists in adding the logarithmic mapping for each level in the scale-space in the original SIFT detection framework. The Difference-of-Gaussians step subtracts the log-mapped levels, making it equivalent to ratio-based edge detection. No additional changes are needed in the SIFT algorithm, making existing implementations easy to be adapted.

IV. Experimental Setup

In order to evaluate the proposed ratio-based SIFT detection method, synthetic images containing geometric forms and real images from the Phos dataset [1] were used as inputs, and results were compared to the original difference-based SIFT detection method. The experiments were divided in three phases: (1) using synthetic images, (2) using images from the Phos dataset with uniform illumination changes and (3) using images from the Phos dataset with non-uniform illumination changes.

The parameters used were based on the ones from the original SIFT implementation [6], and they were the same in both approaches. This means that the input image was subject to an initial blur of $\sigma = 0.5$, and the first Gaussian filter width was $\sigma = 1.6$. From the input image, $s + 3$ images were produced for each octave using a constant $k = 2^{\frac{1}{s}}$ and $s = 3$ as number of scale samples. The contrast threshold discards all extrema in scale-space with normalized contrast less than 0.03 and a curvature ratio of $r = 10$, which eliminates keypoints that have principal curvature ratios greater than 10.

Synthetic images were generated for the initial experiments, with 513x513 pixels in size for better downsampling stability. For the experiments using synthetic images, the object's brightness was decreased until interest points could not be detected by any of the methods, and then it was possible to verify which technique is able to detect keypoints with the lowest contrast.

For the experiments with real world objects, the Phos dataset was chosen for its uniform and non-uniform illumination changes in fifteen different scenes (Phos is one of the very few publicly available datasets with controlled illumination changes). Figure 4 shows the first scene from the Phos dataset with all its illumination scenarios. The different uniform illumination configurations were obtained by adjusting the exposure of the camera between -4 and +4 points from the correct exposure setting. The non-uniform illumination configurations were accomplished with a fixed directional light source and the addition of six levels of uniform illumination [1].

For performance evaluation, the repeatability score method presented by Mikolajczyk and colleagues [10] was used. The repeatability score is computed for a given pair of images as the ratio between the number of region-to-region correspondences and the smallest number of regions among the pair of images. Two regions are considered correspondent if the overlap error between their elliptic regions is sufficiently small. An overlap error of 40% was considered in all experiments, as in the original work by Mikolajczyk and colleagues [10]. Another metric that was used is the number of correspondences, which is computed using the repeatability score and the total number of keypoints detected for each scene.

V. Experimental Results

The results obtained for the proposed ratio-based method were compared to the results obtained for the standard linear SIFT keypoint detection algorithm [5], [6] using synthetic images and the Phos dataset [1] in three experiments.

A. *First Experiment (Synthetic Images)*

In order to test if Wallach's lightness constancy theory [4] would lead to more stable SIFT keypoint detection with respect to illumination changes, the first experiment was based on synthetic images with different contrasts between objects and background.

Figure 5 shows the comparison between the results obtained with the linear (difference-based) and logarithmic (ratio-based) algorithms using $N = 128$. It is clearly noticed that the ratio-based method is capable of identifying regions with lower contrast than the original difference-based implementation. The dimmest circle detected by the linear SIFT has an intensity value of 24, while the dimmest detected by the logarithmic SIFT has an intensity of 2. The last circle (bottom right in the synthetic image shown in Figure 5) has intensity 0, therefore it cannot be detected by either of the methods. The logarithmic method also identifies a larger number of dark blobs surrounding the circles, in multiple scales, while the linear method misses these blobs between the two lower rows of circles.

This is a very promising result which shows that the proposed method is very sensitive even in low contrast regions,

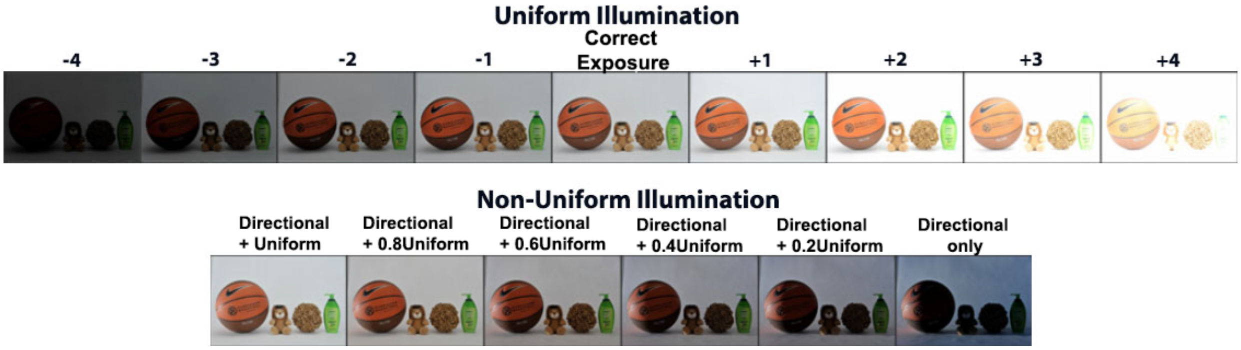
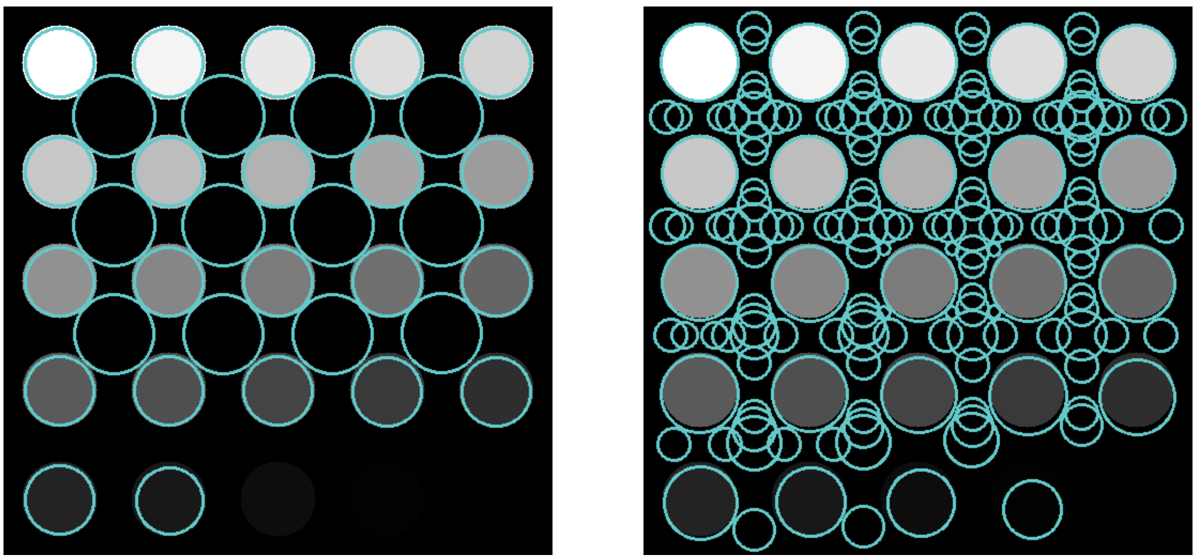


Fig. 4. First scene of the Phos dataset, showing all lighting configurations in the dataset (adapted from [1]). The top row shows uniform illumination, while the bottom row shows non-uniform illumination conditions.



(a) Difference-Based

(b) Ratio-Based

Fig. 5. Comparison between: (a) difference-based and (b) ratio-based (N = 128) SIFT. The blue circles indicate the detected regions and their sizes. The ratio-based algorithm is able to detect regions with an intensity value as low as 2 in a [0,255] range (8-bit intensity values), while the difference-based detects intensities as low as 24.

as expected. The next experiment uses real world scenes with varying illumination conditions, given the solid performance results obtained in this first experiment based on Wallach's theory [4].

B. Second Experiment (Real Images with Uniform Illumination)

The second experiment was based on the Phos dataset [1]. The uniform illumination subset of images was initially chosen in order to provide a stable controlled environment for the initial experiments reported here. Consequently, the subset used in the second experiment is composed by the underexposed and overexposed images from each of the 15 different scenes of the Phos dataset (see top row of Figure 4).

All scenes have their own baseline image, which was acquired with the correct exposure setting (as shown in the center of the top row of Figure 4). Therefore, these baseline images were used as references to compare the performance of the algorithms when the remaining images with illumination changes were used as inputs, as required by the repeatability method by Mikolajczyk and colleagues [10].

Figure 6 shows the average repeatability score (average of all 15 scenes for each exposure configuration) for the studied cases. The linear case represents the difference-based approach while the logarithmic cases represent the ratio-based approach with non-linear mappings. For the underexposed images, the repeatability score performance for the proposed method with N = {2, 4, 8} is worse than the original method. On the other hand, for N = {16, 32, 64, 128, 256} the repeatability score tends to be better for the proposed method, although not significantly higher. For the overexposed images, as the base value N increases, the repeatability score gets significantly higher than the one obtained for the difference-based method. This result can be understood better with Figure 7, which shows the detected keypoints in uniform lighting conditions for the first scene using N = 128 (best overall repeatability

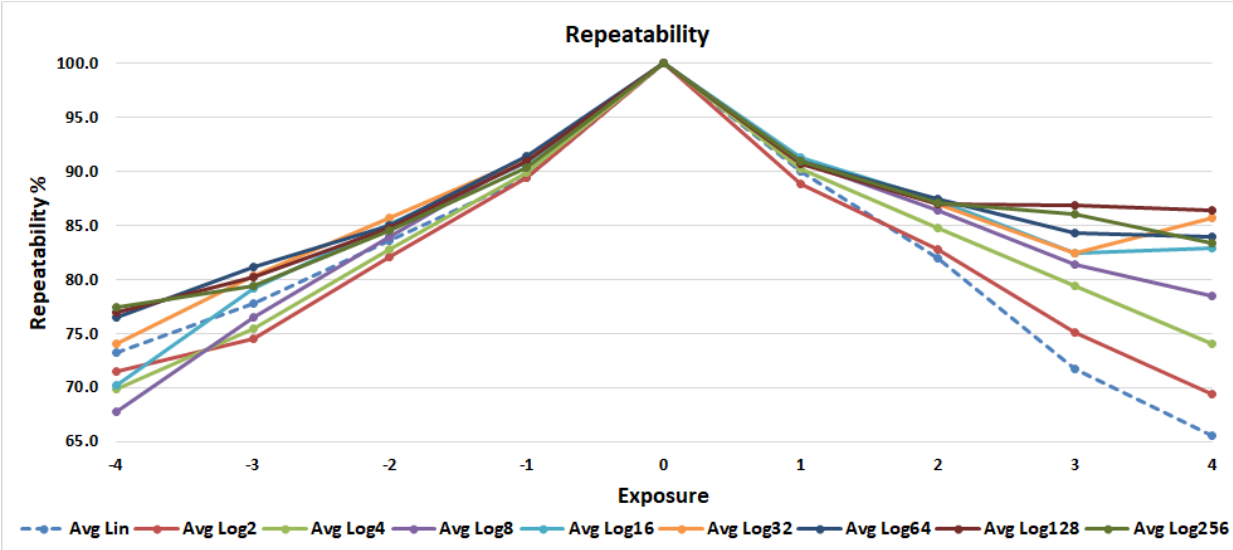


Fig. 6. Comparison of the average repeatability score for the difference-based (Avg Lin) and the ratio-based (Avg Log) approaches for all considered logarithmic bases (N values) considering uniform illumination. The graph shows that repeatability of keypoints is significantly increased for overexposed images when using the logarithmic approach, while the same can not be said regarding underexposed images.

score among underexposed and overexposed images).

Figures 7b and 7e show the keypoints detected for the baseline image of the first scene of the dataset, in which it can be noticed that the linear approach (Figure 7b) detects more keypoints than the logarithmic approach (Figure 7e).

However, our interest here is to investigate how robust these keypoints are with respect to illumination changes. Figures 7a and 7d show the keypoints detected for the minimum exposure setting (-4 points) using linear and logarithmic algorithms, respectively, while Figures 7c and 7f show the keypoints detected for the maximum exposure setting (+4 points).

As can be noticed in Figures 7a and 7d, for underexposed images, the logarithmic approach is capable of detecting more keypoints than the linear approach. The linear approach is still able to identify keypoints, specially in higher levels of the scale-space, meaning that it loses the ability to detect smaller details when compared to the baseline image.

Regarding overexposed images (Figures 7c and 7f), the linear approach has a considerable advantage over the logarithmic approach, detecting interest points from all objects, even with the green object on the right of the scene being barely distinguishable from the background. The ratio-based approach was able to detect just the points that exhibited higher contrast.

It is clear, however, that despite detecting less keypoints in overexposed images, the ratio-based approach reaches better results in terms of repeatability score than the original difference-based implementation. This means that the detected keypoints, lesser in number as they may be, are in the same locations in scale-space as they were for the baseline image.

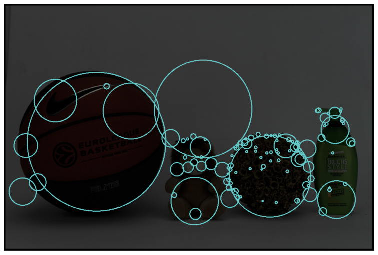
At the same time, although the linear algorithm is able to detect more keypoints, in average, their locations in scale-space differ from the ones detected in the baseline image, which means that they can be considered effectively as new keypoints which are unlikely to be adequate for matching later on, for example.

This discussion raises another important aspect to be considered, which is the number of correct correspondences between the baseline image and its counterparts with illumination changes. Therefore, Figure 8 shows the normalized average number of correspondences between every exposure configuration and the respective baseline image of the scene.

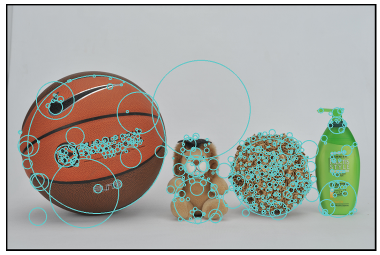
As can be seen in Figure 8, the number of correspondences is significantly larger for the underexposed images when the logarithmic base value increases. On the other hand, despite having a larger repeatability score than the linear mapping, the ratio-based approach detect less keypoints in the overexposed images, which results in less correspondences.

According to the results presented, the ratio-based approach leads to better performance of the SIFT detection algorithm, especially when underexposed images are concerned. Although repeatability scores were not significantly different than the ones obtained for the difference-based approach, the relative number of repeatable keypoints in scale-space was significantly larger.

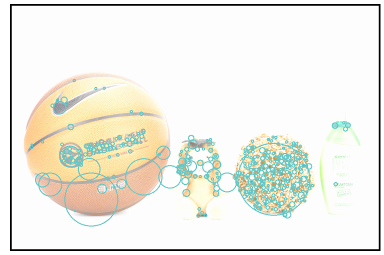
Despite the fact that the ratio-based approach shows worse performance than the difference-based for overexposed illumination, this situation can be avoided using the camera auto-iris setting, which controls how much light reaches the imaging sensor during the acquisition. On the other hand, for image acquisition in dim illumination scenarios, higher ISO values could also be used. However, images acquired with higher ISO values result in too much image noise – the use of the ratio-based approach is particularly interesting for improved SIFT keypoint detection without the need of resorting to noisy image acquisition with higher ISO values.



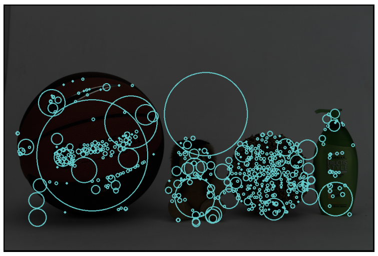
(a) Difference-Based: -4 points



(b) Difference-Based: Baseline



(c) Difference-Based: +4 points



(d) Ratio-Based: -4 points

(e) Ratio-Based: Baseline

(f) Ratio-Based: +4 points

Fig. 7. First scene keypoint detection comparison for N = 128 and extreme exposure configurations. The difference-based approach is displayed in the upper row (a,b and c), while the ratio-based approach is displayed in the lower row (d, e and f). In this example, one can notice how the ability to detect keypoints is reversed when using the proposed approach – the number of detected keypoints is increased for underexposed images, while it is decreased for overexposed ones when compared to the original SIFT approach.

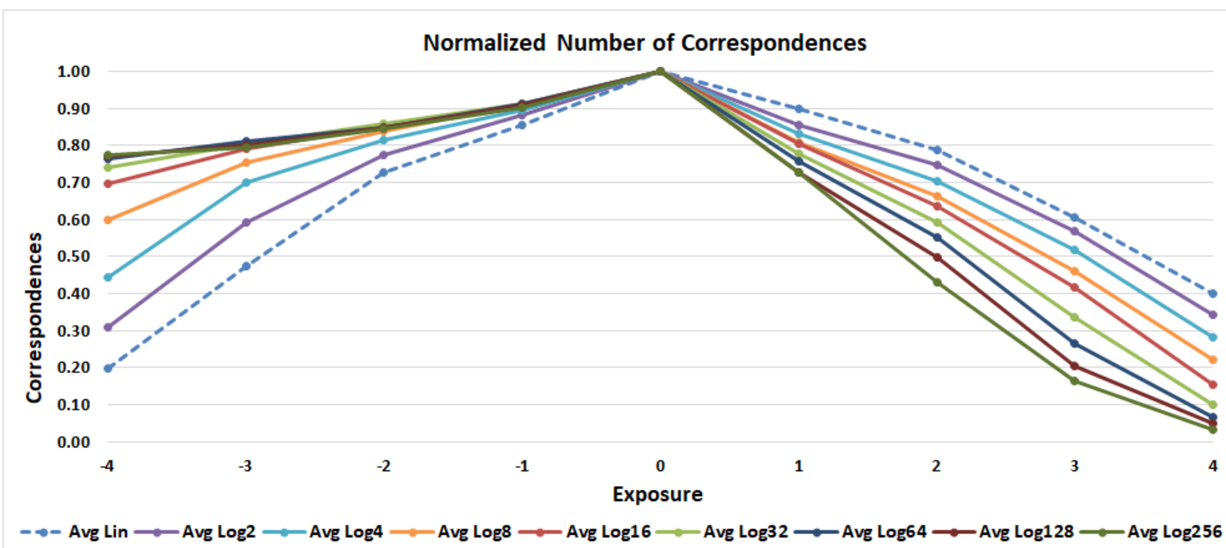


Fig. 8. Comparison of the average number of correspondences for the difference-based (Avg Lin) and the ratio-based (Avg Log) approaches for the uniform illumination scenarios. Values are normalized with respect to the number of keypoints detected in the baseline image.

C. Third Experiment (Real Images with Non-Uniform Illumination)

The most challenging subset of images from the Phos dataset – the ones with non-uniform illumination conditions – were left for the last experiment. These images were acquired with a strong directional light source around with the addition of decreasing uniform illumination in six levels [1].

As can be seen in Figure 9, differently from the experiments with uniform illumination conditions, the proposed ratio-based approach showed better performance regarding repeatability of keypoints in all non-uniform illumination scenarios but one, when compared to the original difference-based approach. This means that the ratio-based approach provides more stable keypoints even with a strong directional light shed upon some objects and shadows cast upon others in the scene.

Figure 10 shows the normalized number of correspondences for the logarithmic approach (for all N values) and for the original difference-based implementation. As can be noticed, the original approach has a slight advantage in the first case, with full uniform illumination intensity. On the other hand,

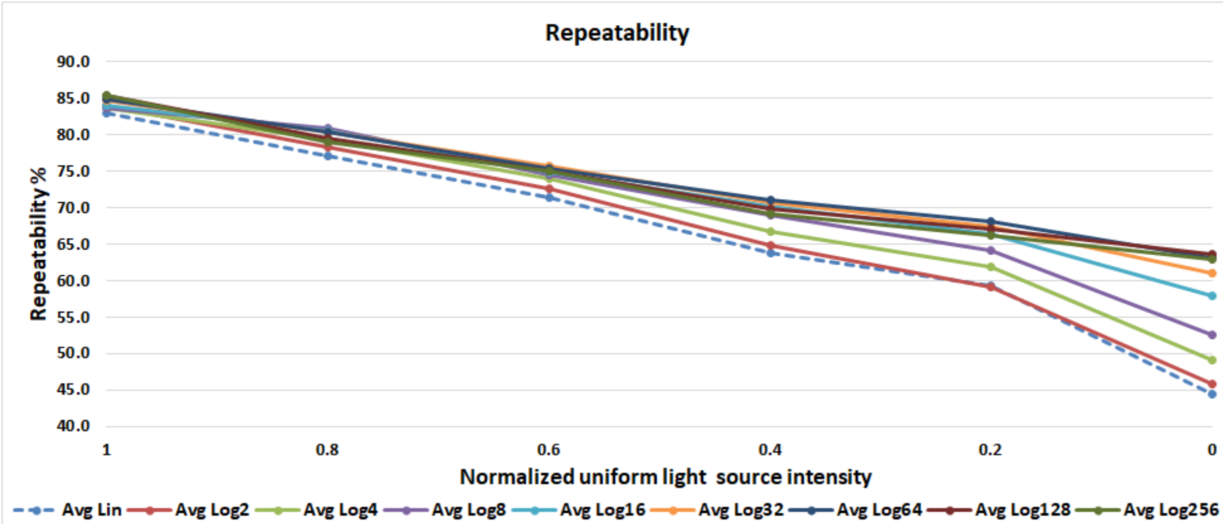


Fig. 9. Comparison of the average repeatability score for the difference-based (Avg Lin) and the ratio-based (Avg Log) approaches considering non-uniform illumination. The graph shows that repeatability score of the proposed ratio-based approach is better in vast majority of cases when compared to the original difference-based approach.

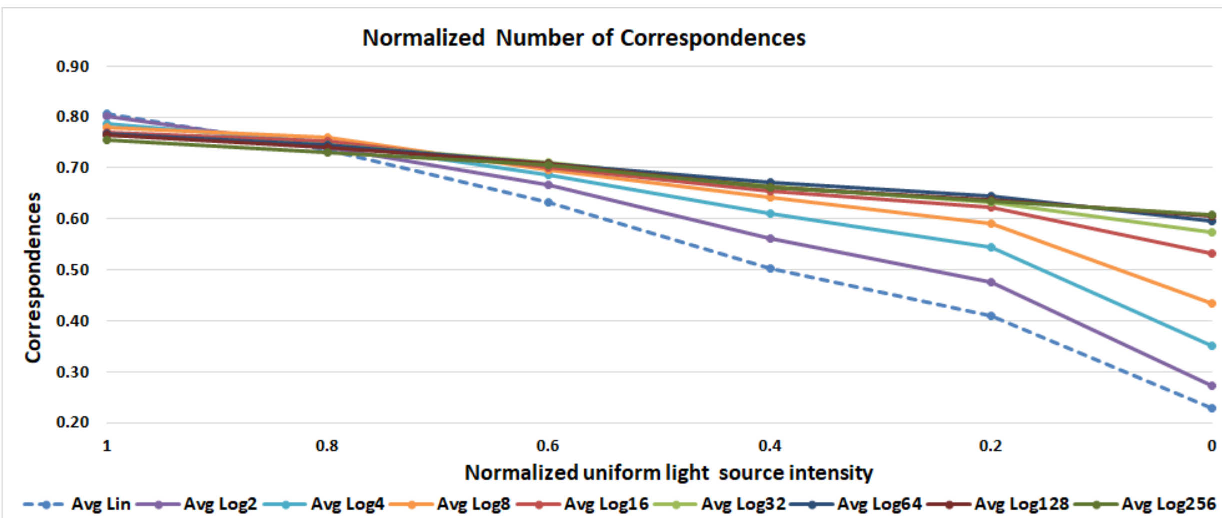


Fig. 10. Comparison of the average number of correspondences for the difference-based (Avg Lin) and the ratio-based (Avg Log) approaches for non-uniform illumination scenarios. Except the first illumination scenario (maximum uniform light source intensity), the remaining experimental results show a clear advantage for the proposed ratio-based implementation.

for the remaining configurations the proposed approach tends to perform increasingly better against the difference-based approach, as the uniform illumination source magnitude is decreased. Regarding number of keypoint correspondences, results for the most challenging illumination setup (just the strong directional light source with no extra uniform illumination source) shows a 35% difference in performance favoring the proposed ratio-based method.

The results obtained in the third experiment show that the proposed ratio-based approach is not as affected by a strong directional light source as the difference-based approach. The non-uniform illumination setup of the Phos dataset can be considered similar to illumination conditions that may be found in outdoor operation (uncontrolled illumination), with multiple light sources and shadows. It is therefore arguable that the proposed approach may perform better than the original SIFT in real world illumination scenarios, making it more suitable for outdoor operation.

VI. Conclusion

SIFT uses a well-known keypoint detector algorithm based on luminance differences that provide some level of illumination invariance. However, according to Wallach [4], contrast perception in natural vision is related to luminance ratios, not differences as usually adopted in artificial vision algorithms. So, in this work we investigated if a ratio-based adaptation in the original SIFT keypoint detector could lead to better performance when subject to changes in illumination.

By using a well-known logarithmic property, it is simple and effective to implement the proposed adaptation in the SIFT scale-space construction. The performance of the proposed method was compared to the original algorithm using both synthetic and real world images with varying uniform and non-uniform illumination setups from the Phos dataset [1]. As evaluation criteria, the repeatability score and the number of correspondences of detected keypoints were used.

For the experiments regarding uniform illumination, the proposed ratio-based approach outperformed the original one for underexposed images, yielding a significantly larger number of stable keypoints, as expected. On the other hand, overexposed images showed a potential weakness of the proposed method, as very few stable keypoints were detected in this case. However, overexposition is thought to be relatively easy to overcome by using automatic exposure settings during image acquisition (e.g. camera auto-iris), while underexposition seems not to be solved as simply as that due to higher noise levels when acquisition sensitivity is increased (e.g. camera auto-ISO).

For non-uniform illumination scenarios, the proposed ratio-based approach showed to be even more valuable. By adding a directional light source, this illumination setup cast shadows over parts of the scene, simulating real world uncontrolled illumination conditions. The results showed significant advantages for the proposed method, with more than 20% increase in repeatability and more than 35% increase in number of correspondences for the most challenging illumination setup (strong directional light source with no extra uniform illumination source).

Overall, the ratio-based approach was able to detect more stable keypoints, even when lesser in absolute numbers when compared to the difference-based approach. The ability to detect stable keypoints in underexposed images and in non-uniform lighting conditions is highly desirable for operation in uncontrolled illumination scenarios, without the need to resort to increased acquisition sensitivity which in turn leads to increased image noise. Future work includes investigating a self-tuning algorithm that can provide an optimal logarithmic base (N) automatically for every illumination scenario and experiments with different algorithms for keypoint detection other than SIFT.

Acknowledgment

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] V. Vonikakis, D. Chrysostomou, R. Kouskouridas, and A. Gasteratos, “A biologically inspired scale-space for illumination invariant feature detection,” *Measurement Science and Technology*, vol. 24, no. 7, pp. 74 024 – 74 036, 2013.

[2] H. Vieira Neto, “Lightness constancy in edge detection: A simple approach using luminance ratios,” in *X Workshop de Visão Computacional*, 2014, pp. 264 – 269.

[3] S. E. Palmer, *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press, 1999.

[4] H. Wallach, “Brightness constancy and the nature of achromatic colors,” *Journal of Experimental Psychology*, vol. 38, no. 3, pp. 310 – 324, 1948.

[5] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Seventh IEEE International Conference on Computer Vision*, 1999, pp. 1150–1157.

[6] ——, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91 – 110, 2004.

[7] A. Jacobsen and A. Gilchrist, “The ratio principle holds over a million-to-one range of illumination,” *Perception & Psychophysics*, vol. 43, no. 1, pp. 1 – 6, 1988.

[8] K. Boonsivanon and A. Meesomboon, “IKDSIFT: An improved keypoint detection algorithm based-on SIFT approach for non-uniform illumination,” *Procedia Computer Science*, vol. 86, pp. 269 – 272, 2016.

[9] M. Huang, Z. Mu, H. Zeng, and H. Huang, “A novel approach for interest point detection via Laplacian-of-Bilateral filter,” *Journal of Sensors*, vol. 2015, 2015.

[10] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, “A comparison of affine region detectors,” *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43 – 72, 2005.