

Optimizing Tone Mapping Operators for Keypoint Detection under Illumination Changes

Aakanksha Rana, Giuseppe Valenzise, Frederic Dufaux

► **To cite this version:**

Aakanksha Rana, Giuseppe Valenzise, Frederic Dufaux. Optimizing Tone Mapping Operators for Keypoint Detection under Illumination Changes. 2016 IEEE Workshop on Multimedia Signal Processing (MMSP 2016), Sep 2016, Montréal, Canada. 2016 IEEE Workshop on Multimedia Signal Processing (MMSP 2016), 2016, <<http://mmsp2016.ece.mcgill.ca/default.html>>. <hal-01349708>

HAL Id: hal-01349708

<https://hal-imt.archives-ouvertes.fr/hal-01349708>

Submitted on 28 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimizing Tone Mapping Operators for Keypoint Detection under Illumination Changes

Aakanksha Rana, Giuseppe Valenzise, Frederic Dufaux
LTCI, CNRS, Télécom ParisTech, Université Paris Saclay, 75013 Paris, France
{aakanksha.rana, giuseppe.valenzise, frederic.dufaux}@telecom-paristech.fr

Abstract

Tone mapping operators (TMO) have recently raised interest for their capability to handle illumination changes. However, these TMOs are optimized with respect to perception rather than image analysis tasks like keypoint detection. Moreover, no work has been done to analyze the factors affecting the optimization of TMOs for such tasks. In this paper, we investigate the influence of two factors— Correlation Coefficient (CC) and Repeatability Rate (RR) of the tone mapped images for the optimization of classical Retinex based models to enhance keypoint detection under illumination changes. CC-based optimized models aim at increasing the similarity of the tone mapped images. Conversely, RR-based optimized models quantify the optimal detection performance gains. By considering two simple Retinex based models, i.e., Gaussian and bilateral filtering, we show that estimating as precisely as possible the illumination, CC-based optimized models do not necessarily bring to optimal keypoint detection performance. We conclude that, instead, other criteria specific to RR-based optimized models should be taken into account. Moreover, large gains in performance with respect to existing popular TMOs motivate further research towards optimal tone mapping technique for computer vision applications.

Index Terms

High dynamic range imaging, keypoint detection, tone mapping, Retinex.

1. Introduction

The ability to capture and reproduce detailed information over a large range of radiance values has brought potential interest in high dynamic range (HDR) imagery for several computer vision problems like object tracking, face recognition and privacy protection [1–3]. A key advantage of using this technology is its robustness to drastic illumination changes which has been lately studied in the context of keypoint detection [4–6].

Conventionally, keypoint detection [7], which is the primary step in many computer vision processing pipelines, has been designed and optimized with respect to low dynamic

range (LDR) imagery, generally represented using 8-bit integer pixel values. Conversely, HDR pixels are real-valued and proportional to the physical luminance of the scene, expressed in cd/m^2 . As a result, in order to employ conventional keypoint detection schemes, one feasible solution is to convert HDR images in a suitable 8-bit LDR representation. This is achieved through *tone mapping operators* (TMO) [8]. Traditionally, TMOs have been designed to display HDR content on LDR displays. As a consequence, they mainly aim at reproducing scenes which appear natural and visually pleasing [9,10].

Recent studies [4,6] reveal that visually optimized TMOs do not perform consistently well for keypoint detection in the presence of challenging transformations such as scaling, rotation or illumination. Indeed, in image analysis tasks the performance criteria are not related to visual appearance but to task-specific measures like keypoint repeatability rate [7]. However, there are some TMOs which perform better than the others.

A potential class of TMOs includes Retinex based approaches [11,12]. A Retinex based model normalizes the estimated luminance from the given HDR radiance map, thus the resulting reflectance image is assumed to retain the fine details present in the scene. However, Retinex is a mathematically ill-posed problem [13]. As a consequence, the obtained reflectance maps are not exact. Furthermore, these perceptually optimized Retinex [11,12] based TMOs aim at finding a visually pleasing reflectance map (considering certain luminance assumptions) for a single scene. On the contrary, keypoint detection performance is evaluated on test-reference image pairs. Therefore, it is evident that Retinex based tone mapping models need to be optimized for efficient keypoint detection. Above all, there is no related work in the literature aiming at designing or comprehending the criteria involved for a detection-optimized tone mapping technique.

In light of these observations, this paper targets two major questions: if the Retinex based models are optimized for keypoint detection in lighting change scenario, what are the influential factors for optimization, and what are the corresponding gains.

Motivated by these questions, we propose to investigate the optimization of Retinex based models using two factors: Correlation Coefficient (CC) and Repeatability Rate (RR) of reflectance images obtained as a result of Retinex based TMs. In addition, we propose to employ the image pairwise optimization strategy, contrary to single image tuning of traditional tone mapping techniques.

The work presented in this document was supported by BPIFrance and Région Ile de France, in the framework of the FUI 18 Plein Phare project.

The CC measures the statistical similarity between a pair of reflectance maps. Theoretically, a CC-optimized Retinex based TMOs should improve keypoint detection performance as a high similarity of reflectance image pairs should increase the probability of detection of similar keypoints. In simple words, the CC-optimized models help analyzing the relationship between statistical correlation of the reflectance image and the keypoint detection performance. On the other hand, our second optimization factor RR is a conventional performance measure of keypoint detection algorithms. Hence, we argue that optimization of a TMO with respect to the corresponding RR could estimate the optimal detection performance gains. RR is computed on repeated occurrences of detected keypoints in test and reference images.

To evaluate our optimization methods, we consider two popular Retinex based Gaussian [11] and Bilateral [12] models. We choose these models mainly because they are simple (less parametric dependency and easy to tune), most widely used and less computationally complex. In addition, we test our optimized approaches using HDR image dataset with illumination changes datasets [4] and two classical corner and blob detection schemes (Harris and SURF) for keypoint detection. Undoubtedly, there are many Retinex based models and keypoint detection techniques in the literature, but they are beyond the scope of this paper. We mainly focus on the factors for optimizing TMOs that can impact the keypoint detector performances.

In this paper, we optimize both considered Retinex models with respect to CC and RR and record the performance for the task of keypoint detection. Henceforth, we address the relationship between each of the optimized models and the keypoint detection performance which is essential to redefine the modeling strategy needed for the keypoint detection-optimized tone mapping technique. In addition, we explicitly compare different existing local and global high performing TMOs, as provided in previous studies [4,6], with our RR based optimized tone mapping models and record substantial gains with our RR-optimized models.

The paper is organized as follows. In Section 2 we provide the details of the evaluation framework. We present the experimental results and analysis in Section 3. Finally, the conclusions are drawn in Section 4, along with future research directions.

2. Evaluation Framework

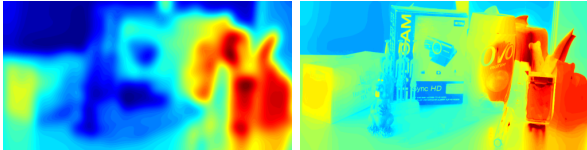
In this section, we first detail the considered Retinex based models used in our evaluation. Next, we briefly discuss the considered feature detection methods, followed by metrics and dataset selection. Finally, we detail the optimization strategies of considered models.

2.1. Considered Retinex Based Models

Retinex based algorithms have been used for several applications including dynamic range compression. However, the goal has been limited to visual perception [11,12]. According to Retinex theory of physical image modeling,



(a) Original (I)



(b) Gaussian (L_g)

(c) Bilateral (L_b)



(d) Gaussian Reflectance (R_g) (e) Bilateral Reflectance (R_b)

Figure 1: Reflectance images R_g and R_b from original image I using the Gaussian and Bilateral luminance maps L_g and L_b respectively.

we assume that I , the HDR image to be tone mapped, is the product of the luminance L of the scene (which varies with different illumination conditions) and of the reflectance R characterizing objects of the scene, i.e., $I = R \cdot L$. The luminance L is generally assumed to be spatially smooth [14], while reflectance contains fine-grained details, texture and edges which are relevant for detection [7,15]. Nevertheless, estimating L from a single image I is a well-known ill-posed and computationally expensive problem [13], which has been studied for many decades [16]. Once L is estimated, the final reflectance image is given by $R = I/L$.

In this paper, we employ two well-known approaches using: a) Gaussian model [11] and b) Bilateral model [17] for tone mapping. In the following, we briefly describe these luminance estimation models for our TM technique:

a). **Gaussian tone mapping** (GTM) model gives the reflectance image R_g as $R_g = I/L_g$ where

$$L_g = I * G_\sigma, \quad (1)$$

where G_σ is a Gaussian filter where the parameter σ depends on image size $[m \times n]$, i.e. $\sigma = \alpha \cdot \max(m, n)$. When targeting visual perception, the parameter σ is tuned so as to reduce visual artifacts like halos observed around the detected edges. This model with single parameter is simple and computationally very fast.

b). **Bilateral tone mapping** (BTM) model is a a precise, non-linear and edge preserving filter where R_b is computed

as $R_b = I/L_b$. The luminance estimation $L_b(x)$ is given as:

$$L_b(x) = \frac{1}{N} \sum_{y \in S} G_{\sigma_s}(\|x - y\|) \cdot G_{\sigma_r}(\|I_x - I_y\|) I_y, \quad (2)$$

where x and y are pixel locations, S is the set of neighborhood locations, G_{σ_r} and G_{σ_s} are Gaussian filters with variances σ_r and σ_s referred to as range and spatial parameter respectively. N is a normalization factor :

$$N = \sum_{y \in S} G_{\sigma_s}(\|x - y\|) G_{\sigma_r}(\|I_x - I_y\|). \quad (3)$$

It is important to note here that when the range parameter increases, the model gradually approaches Gaussian convolution. This is mainly because the Gaussian G_{σ_r} widens and flattens, and essentially, it becomes nearly constant over the intensity interval of the image. Conversely, when the spatial parameter increases, the larger details like edges get smoothed in the image.

An example of luminance estimation with their corresponding reflectance image is shown in Figure 1.

2.2. Keypoint point detection

Keypoint detection has been widely studied in computer vision literature where several techniques have been proposed and evaluated [7] taking into account different challenging transformations. In this paper, we focus on the two most widely used keypoint detection schemes, i.e., *corner* and *blob* detectors. We select two common detectors that have been used in previous HDR imagery based evaluations [4,5] and are often used in several real-time applications. For corner interest point detector, we employ the popular *Harris corner point* detector [18], which is based on the autocorrelation score computed from local intensity change in an image. For blob detection, we carried out experiments with the highly robust *SURF*[19] detector. We used MATLAB’s Computer Vision System Toolbox for both Harris and SURF keypoint detection, where we can select the strongest n keypoints based on their strengths.

2.3. Metrics

In this study, we build our contribution using following metrics.

- **Repeatability Rate (RR)** is a standardized method detailed in [7] to measure the detector accuracy. It is given as the fraction of keypoints detected in the reference image which are repeated in the test image to the minimum of a total number of detected points in test or reference image. A keypoint is considered to be repeated in the test image if: a) it is detected as a keypoint in the test image, and b) it lies in a circle of radius ϵ centered on the projection of the reference keypoint onto the test image. ϵ determines the keypoint detection error rate. RR is given as $\frac{Q_i(\epsilon)}{\min(n_r, n_i)}$, where $Q_i(\epsilon)$ is the number of keypoints detected in the reference image which are repeated in the test image, n_r, n_i is

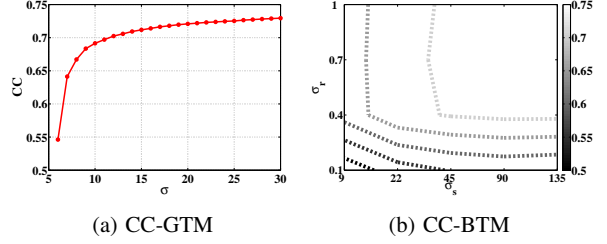


Figure 2: *Parameters vs Correlation Coefficient (CC) for Project Room dataset.* (a) σ vs CC for Gaussian tone mapping (GTM) model. (b) σ_r and σ_s contours for Bilateral tone mapping (BTM) model with color magnitudes showing average CC scores.

the number of detected keypoints in reference and test image respectively.

- **Correlation Coefficient (CC)** is well-known to quantify the strength of a linear relationship between two variables. In this study, CC metric measures the correlation between two image maps. Value close to 1 indicates a positive linear relationship and 0 corresponds to no relationship between image maps.

2.4. Datasets

We considered the HDR dataset with substantial illumination changes proposed in our previous work [4]. It is composed of 2 parts: Project-Room with 8 lighting conditions and Light-Room with 7 lighting conditions. The Project Room dataset is more challenging in terms of lighting transformations, textured objects and stark shadows whereas Light Room dataset is comparatively simpler with less challenging illumination conditions.

2.5. Optimization of Retinex models

Let $I_1(x)$ and $I_2(x)$ be two images of the same scene which is illuminated by two different illumination maps $L_1(x)$ and $L_2(x)$. According to Retinex theory $I_1(x) = L_1(x) \cdot R_1(x)$ and $I_2(x) = L_2(x) \cdot R_2(x)$. An ideal Retinex algorithm estimates $L_1(x)$ and $L_2(x)$ such that two corresponding reflectance maps are equal $R_1(x) = R_2(x)$. In such an ideal scenario, keypoint detection performance should be enhanced considerably as the keypoints in reflectance maps would be identical.

However, Retinex is a mathematically ill-posed problem [13]. In practice, it often implies that $R_1 \neq R_2$. Besides this, all perceptually optimized Retinex based models aim at finding the best luminance estimation for a given scene such as $I_1 = L_1 \cdot R_1$, rather than optimizing on aforementioned image pairs like (R_1, R_2) which is the classical way of measuring keypoint detection performance.

Therefore, we firstly investigate the maximization of the correlation between reflectance image pairs. The motivation is that highly correlated reflectance maps should result in detected keypoints that are alike, thereby, enhancing the keypoint repeatability. Alternatively, we also investigate the

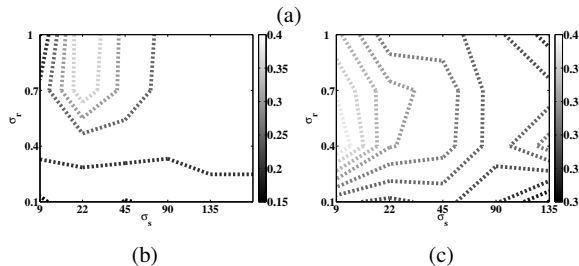
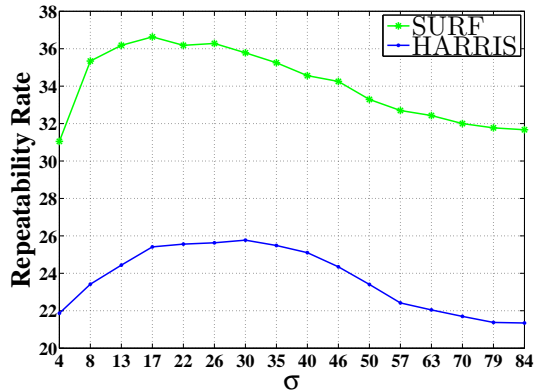


Figure 3: *Parameters vs Repeatability Rate (RR) for Project Room dataset.* (a) σ vs RR for both SURF and Harris detector for repeatability rate Gaussian TM (RRGTM). (b) σ_r and σ_s contours for Harris detector and (c) for SURF detector for repeatability rate Bilateral TM (RRBTM) with color magnitudes showing RR scores.

optimization of the considered models with respect to repeatability rate which will help to analyze the maximum gains achievable with conventional TMOs.

In summary, we optimize the considered Retinex based GTM and BTM models in two ways: 1) with respect to correlation of reflectance maps of image pairs and, 2) with respect to detector repeatability. For the first method, we iteratively optimize the GTM and BTM models parameters with respect to CC on both datasets using each detector. More specifically, we iteratively tune the parameter σ for GTM and σ_s , σ_r for BTM with the aim of maximizing the overall CC using each detector.

Correlation based parameter tuning for Project Room dataset illustrated in Figure 2 (a). It depicts that for GTM, a higher σ , i.e. high variance Gaussian blur, minimizes the absolute differences between the reflectance image pairs. The same observation also holds for range and spatial parameters of BTM model as shown in Figure 2 (b). Thereafter, using these correlation based optimized models, we generate the tone mapped images as correlation-coefficient-Gaussian-tone-mapping (CCGTM) and correlation-coefficient-Bilateral-tone-mapping (CCBTM) for both datasets.

For the second method of optimization with respect to detector repeatability, we tune the GTM and BTM parameters with the aim of maximizing the overall repeatability rate (RR) of keypoints using both Harris and SURF detec-

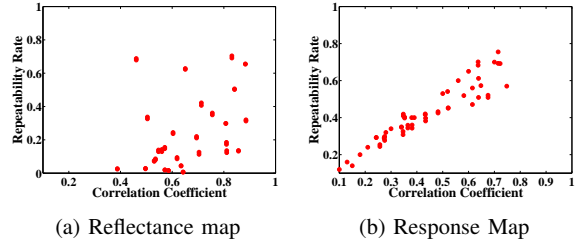


Figure 5: *Scatter plots.* (a) correlation coefficients of reflectance maps $CC(R_i, R_j)$ vs corresponding repeatability rate $RR(R_i, R_j)$, (b) correlation coefficients of response maps $CC(Resp_m, Resp_n)$ vs corresponding repeatability rate $RR(R_m, R_n)$ for HDR log-encoded Project room dataset.

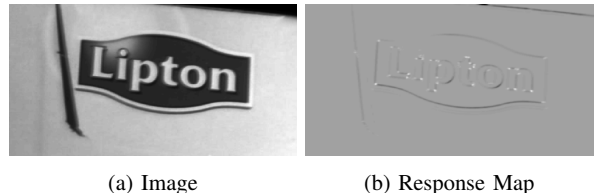


Figure 6: An example showing (a) image and its corresponding (b) Harris response map.

tor. Corresponding results are shown in Figure 3 (a) for the Gaussian model, and in Figure 3 (b) and (c) for the Bilateral model. Similarly, we generate the tone mapped images, repeatability-rate-Gaussian-tone-mapping (RRGTM) and repeatability-rate-Bilateral-tone-mapping (RRBTM), for both datasets and using each detector.

3. Experimental Results and Discussion

We evaluate the CCGTM, CCBTM, RRGTM, RRBTM optimized tone mappings models for keypoint detection using Harris and SURF detectors in Figure 4(a),(b). Additionally, we compare 5 different local and global high performing TMOs [9–12,20] with our RRGTM and RRBTM tone mapping models as shown in Figure 4(c),(d).

For each TM, we measure the overall keypoint detection accuracy using the RR performance metrics. Initially, we compute the individual RR using a particular detector over all the possible images pairs for each scene (Light Room and Project Room), i.e., each lighting condition of a scene is, in turn, the reference and the other conditions of that scene are the test images. Thereafter, we compute the average of RR over all such possible image pairs (52 pairs for Project Room and 42 pairs for Light Room) for a particular scene. Similar to [4,6], our keypoint detection scheme is confined to the strongest 1000 keypoints. This is mainly to ensure a fair comparison of the blob or corner keypoint detection, as different detectors result in a highly different number of keypoints. We use a fixed detection error rate, i.e. $\epsilon = 3px$, which is .03% of image size (see section 2.3).

CC vs RR based optimization. We observe from Figure 4(a),(b) that CCGTM and CCBTM records substantial gap in performance with respect to RRGTM and RRBTM

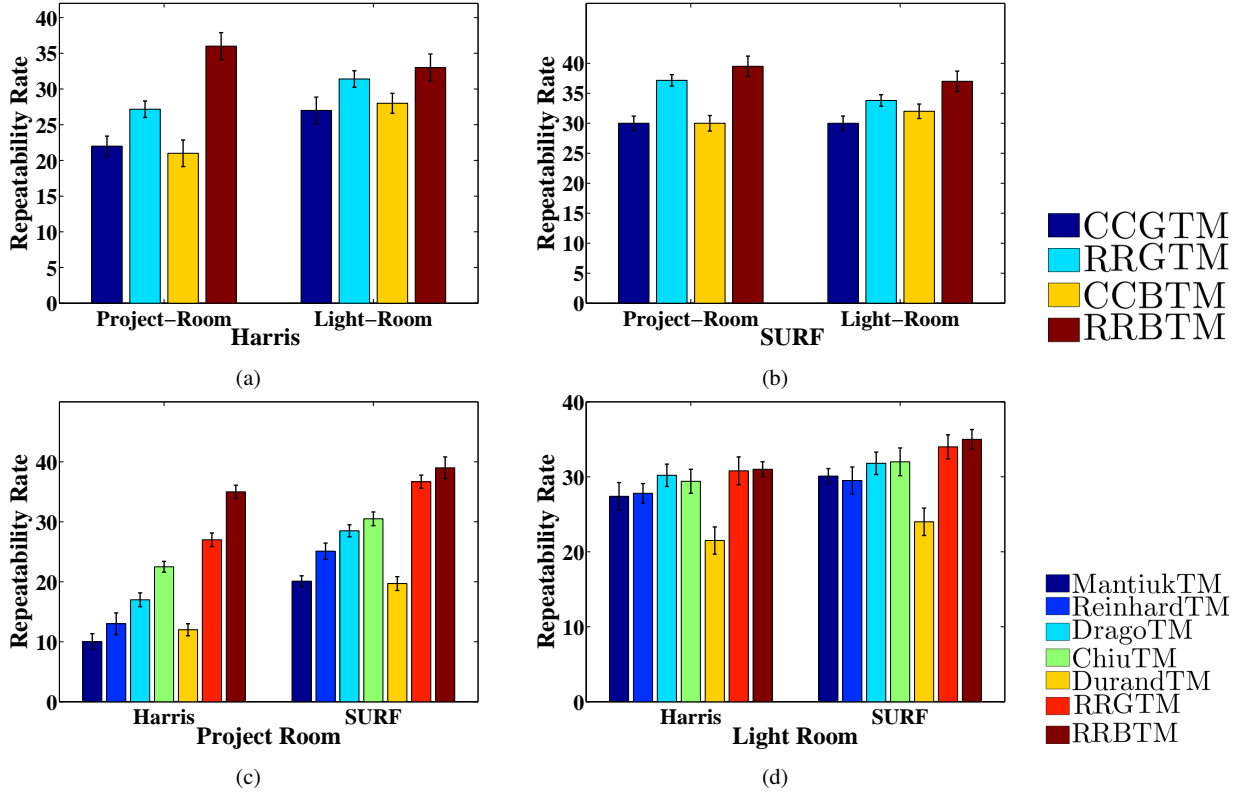


Figure 4: Row 1. (a) and (b) Average repeatability score and standard deviation for both correlation and response based optimized approaches using Harris and SURF detector respectively. Row 2. (c) and (d) Average repeatability score and standard deviation for the reflectance models (GTM and BTM) and other commonly used TMs on Project Room and Light Room dataset

using both Harris and SURF detectors. We can conclude that high correlation between reflectance image pairs does not directly guarantee the stability of keypoint detection. However, this can be explained as follows. Keypoints are selected after thresholding the response map, which depict the probability of a pixel to be detected as a keypoint, generally relying on second order derivatives. An example of response map using Harris detector [18] is shown in Figure 6, where high pixel values indicate higher likelihood to be considered as keypoint. As a consequence, local keypoints are in general sparingly distributed in detailed areas of an image. This process is much more complex than a simple statistical correlation computed at the pixel level. This is further illustrated in Figure 5. It can be observed that correlation of response maps $CC(Resp_m, Resp_n)$ are linearly proportional to RR, whereas, the correlation of the reflectance image maps $CC(R_i, R_j)$ shows scattered behavior. Therefore, this leads to the major conclusion of this study. In order to design a keypoint-detection-optimal tone mapping, the traditional Retinex based approaches need to take into account the detector response maps while estimating the reflectance images for a given scene.

Comparison with Traditional TMs. From Figure 4(c),(d), we show that optimizing the traditional Retinex models in Eqs (1) and (2), RRGTM and RRBTM

respectively, lead to large performance gains in terms of RR when compared to existing local and global TMs. It shows the necessity to optimize the tone mapping operators with respect to detection tasks.

Finally, we also observe that the performance gains are significantly larger for Project Room than Light Room. This is mainly due to the fact that lighting transformations are much smaller for Light Room dataset, which entails less performance variations in RR when comparing different tone mappings.

4. Conclusion

In this paper, we have presented the impact of two factors: Correlation Coefficient (CC) and Repeatability Rate (RR), for the optimization of Retinex based models, aiming at enhancing the keypoint detection performance under drastic lighting change scenarios. The proposed image-pair based optimized models are tested for real HDR lighting change datasets using two classical (Harris and SURF) detectors. The CC-based optimized models maximize the similarity of Retinex-resulting-reflectance images and RR-based optimized models determine the optimal performance gains. Our tests demonstrate the advantage of using optimized models over perceptually motivated models for keypoint de-

tection task. Furthermore, our study suggests that a keypoint detector-based optimized TMO has a lot of potential as CC based optimized models lag behind the RR optimized models by a large margin. To this end, we find that detector response maps (which are the components of keypoint detection schemes) can somehow be exploited to achieve this goal. In future, we plan to extend this work by designing an optimal Retinex based tone mapping operator for keypoint detection, which takes into consideration the detector response maps.

References

- [1] L. Chermak, N. Aouf, and M. Richardson, "HDR imaging for feature tracking in challenging visibility scenes," *Kybernetes*, pp. 1129–1149, 2014.
- [2] P. Korshunov, M. V. Bernardo, A. M. G. Pinheiro, and T. Ebrahimi, "Impact of tone-mapping algorithms on subjective and objective face recognition in hdr images," in *International ACM Workshop on Crowd-sourcing for Multimedia (CrowdMM)*, 2015, number EPFL-CONF-210823.
- [3] M. Rerabek, L. Yuan, L. Krasula, P. Korshunov, K. Fliegel, and T. Ebrahimi, "Evaluation of privacy in high dynamic range video sequences," 2014.
- [4] A. Rana, G. Valenzise, and F. Dufaux, "Evaluation of feature detection in HDR based imaging under changes in illumination conditions," in *IEEE International Symposium on Multimedia, ISM 2015, Miami, USA, December, 2015*, 2015, pp. 289–294.
- [5] P. Bronislav, A. Chalmers, and P. Zemčík, "Feature point detection under extreme lighting conditions," in *Spring Conference on Computer Graphics*, 2012, pp. 156–163.
- [6] P. Bronislav, A. Chalmers, P. Zemčík, L. Hooberman, and M. Čadík, "Evaluation of feature point detection in high dynamic range imagery," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 141 – 160, 2016.
- [7] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vision*, pp. 151–172, June 2000.
- [8] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video: From Acquisition, to Display and Applications*, Academic Press, 2016.
- [9] R. Mantiuk, K. Myszkowski, and H. P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Percept.*, vol. 3, no. 3, pp. 286–308, July 2006.
- [10] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, pp. 267–276, July 2002.
- [11] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, "Spatially nonuniform scaling functions for high contrast images," in *Proceedings of Graphics Interface '93*, Toronto, Ontario, Canada, 1993, GI '93, pp. 245–253.
- [12] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, 2002, SIGGRAPH '02, pp. 257–266.
- [13] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *Image Processing, IEEE Transactions on*, Mar 1997.
- [14] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [15] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [16] B. K. P. Horn, "Determining lightness from an image," *Computer graphics and image processing*, 1974.
- [17] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the Sixth International Conference on Computer Vision*, Washington, DC, USA, 1998, ICCV '98, pp. 839–.
- [18] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.
- [19] H. Bay, T. Tuytelaars, and L. V. Gool, "Surf: Speeded up robust features," in *In ECCV*, pp. 404–417. 2006.
- [20] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, pp. 419–426, 2003.