HAL
open science

Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions

Aakanksha A Rana, Giuseppe Valenzise, Frederic Dufaux

To cite this version:

Aakanksha A Rana, Giuseppe Valenzise, Frederic Dufaux. Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions. IEEE International Symposium on Multimedia (ISM) , Dec 2015, Paris, France. 10.1109/ISM.2015.58 . hal-01331624

HAL Id: hal-01331624

https://hal.science/hal-01331624v1

Submitted on 14 Jun 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.

Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions

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*—High dynamic range (HDR) imaging enables to capture details in both dark and very bright regions of a scene, and is therefore supposed to provide higher robustness to illumination changes than conventional low dynamic range (LDR) imaging in tasks such as visual features extraction. However, it is not clear how much this gain is, and which are the best modalities of using HDR to obtain it. In this paper we evaluate the first block of the visual feature extraction pipeline, i.e., keypoint detection, using both LDR and different HDR-based modalities, when significant illumination changes are present in the scene. To this end, we captured a dataset with two scenes and a wide range of illumination conditions. On these images, we measure how the repeatability of either corner or blob interest points is affected with different LDR/HDR approaches. Our observations confirm the potential of HDR over conventional LDR acquisition. Moreover, extracting features directly from HDR pixel values is more effective than first tonemapping and then extracting features, provided that HDR luminance information is previously encoded to perceptually linear values.**

***Keywords*-High dynamic range imaging, feature extraction, detectors, keypoints, tone mapping.**

I. Introduction

One of the major challenges being faced by several recognition approaches such as visual localization, detection and classification, is the change in appearance of objects captured across a wide range of illumination conditions. In surveillance applications, for instance, adverse lightening conditions can significantly deteriorate the performance of feature detection, which is the first block of the feature extraction pipeline. High dynamic range (HDR) imaging [1] enables to partially overcome these limitations, by acquiring multiple images with different exposure times and combining them in a single HDR picture, thereby preserving details in both dark and bright regions.

Conventionally, feature extraction algorithms [2] have been designed and optimized with respect to low dynamic range (LDR) images, generally represented using an integer 8-bit representation approximately linear to human perception. Conversely, HDR pixels are real valued and proportional to the physical luminance of the scene, expressed in cd/m^2. As a result, it is not clear which LDR feature extraction techniques might be applied to HDR images, and which could be the best modalities to do so.

In this paper we try to provide a comprehensive understanding of the potential of HDR for feature detection. In particular, we consider the following research questions: a) *is HDR capable to achieve substantial gains in terms of feature stability to luminance changes compared to LDR*? b) if yes, *which is the best way to use such HDR images*, i.e., direct real-valued luminance, or HDR converted to LDR format through a tonemapping operation (TMO) in order to be compatible with standard feature extraction techniques?

To answer these questions, a dataset of HDR and LDR images is built, consisting of two setups, each one illuminated with seven and eight different lighting conditions, respectively. This dataset is challenging in terms of texture reflectance of objects, presence of shadows and variety of illumination sources. For each lighting scene, we consider a number of image formats, including linear or perceptually encoded HDR values, the subjectively best LDR exposure, and several local or global tone-mapped pictures. Next, we detect features from each lighting scene, and we compute the repeatability of detected interest points in all the other illumination settings, in order to estimate the average feature stability. This is accomplished using two popular corner point (Harris [3]) and blob detectors (SURF [4]).

With respect to previous work [5] [6], we focus on feature stability under illumination changes along with analyzing the performance of many popular tone-mapping approaches [7]–[9] which have been evaluated thoroughly from a perceptual point of view, but whose effectiveness in feature extraction has not been investigated so far. Finally, we explicitly compare direct feature detection on HDR images with a tonemap-then-extract approach [10], for which we mainly investigated the linear, log and perceptually uniform (PU) encoding of pixel values of the HDR modality.

The remainder of this paper is organized as follows: in the next section we review related work in the domain of HDR image analysis. Section III describes the evaluation set-up, including the feature detectors, tone mapping techniques, datasets and HDR adaptation schemes used. We present the experimental results and discussion for the proposed methodologies in Section IV and the conclusions in Section V.

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.

II. Background and Related Work

One common way to generate HDR images is by capturing multiple LDR pictures of a scene at different exposure times, in order to estimate a signal proportional to the physical luminance of the scene [1]. The luminance information is generally represented using floating point formats that can use up to 32 bits per channel per pixel, differently from traditional 8-bit LDR formats that store gamma-encoded values (approximately linear to perception). Conventional display technology, as well as the majority of computer vision algorithms, assume that the input image is LDR.

In order to compress the dynamic range of a HDR image to LDR, a great variety of *tonemapping* operators (TMO) have been proposed in the past years. These algorithms can be broadly classified into *global* approaches, where the same compression function is applied to all the pixels of the image, and *local* techniques, where a tone-mapped pixel depends on the values of neighboring pixels. The performance of TMO's have been widely studied from a perceptual point of view [11], [12] and, more recently, for specific applications such as video compression [13]. In general, global TMO's can preserve the overall perceived contrast of the original HDR, while local TMO's are better in conveying local structure, i.e., fine details and edges, although in some cases yield unnatural LDR results. Nevertheless, TMO's have been rarely explored for applications such as feature extraction, and little research has been done on finding the best modalities to use the HDR content.

Early work on optimizing TMO for surveillance applications was carried out by [14], who propose to combine the properties of local and global TMO's for object detection and tracking. However, that work lacks the comparison in terms of detection accuracy with other TMO's. The work in [15] presents instead a normalization approach, where TMO is used to remove lighting-dependent information from an HDR picture, and leaving only the object's texture. Results are shown in terms of SIFT descriptors matching performance, on a limited dataset of two images, in comparison to two popular TMO's. However, the first part of feature extraction, i.e., keypoint detection, is still left out from evaluation.

The evaluation of feature extraction on HDR content can be conducted at different levels of the feature processing chain. At the application level, one can measure the performance of tasks such as object detection in outdoor locations, pedestrian or vehicle tracking [6], [16], or privacy protection [17]. For instance, [16] compared the feature tracking performance of SURF and SIFT descriptors using a dataset of both indoor and outdoor HDR images. However, by looking at the results of a specific vision task, it is difficult to draw precise conclusions on what makes certain HDR modalities perform better than others. In this paper, we take a lower level viewpoint and concentrate on the detection of interest points. The authors of [10] report an increase in the number of detected feature points using gradient-based tone-mapped HDR images over LDR. However, the number of detected feature points is not itself a sufficient indicator of detection performance. In our work, we employ instead the standard criterion of feature *repeatability* [2].

(a) Drago GTM [8] (b) Chiu LTM [7] (c) Best exposure LDR

Figure 1: Sample Images from Project Room (Row-1) and Light-Room (Row-2) datasets with local, global tone mappings and best exposures LDR.

The evaluation methodology that is perhaps most related to the one in this paper is the work of Pribyl et al. [5], which presents an evaluation of the repeatability of state-of-the-art key-point detectors on images under different transformations (lighting, viewpoint, distance) for different LDR/HDR modalities, including simple global and local TMO's but *not* the original HDR values. In this work, we corroborate some of the findings of that work, and we extend its analysis by evaluating popular TMO's used for display, as well as feature extraction directly on HDR with either linear or perceptually encoded pixel values.

III. Evaluation Setup

In this section, we describe the proposed image dataset, as well as the choice of feature detectors and HDR/LDR modalities considered in this study.

A. Proposed dataset

Accuracy measurement of feature detection is based on repeatability criterion, which further relies on the precise localization of key-points in both reference and test images, so that correspondences between detected features points can be unequivocally assessed [2]. Unfortunately, the great majority of existing HDR image and video datasets are not adapted to this end, as images are not geometrically calibrated. To the authors' knowledge, the only such existing HDR dataset adequate for a confined low-level evaluation has been proposed in [5], *2D and 3D Lighting Dataset* ,where a scene with controlled lighting conditions has been captured. However, the number of lighting conditions is quite limited. In this paper, we propose two different lighting setups: *Project Room* and *Light Room* (Figure 1), focusing mainly on lighting changes and variation in dynamic range of the scenes, which are recognized to be some of the most

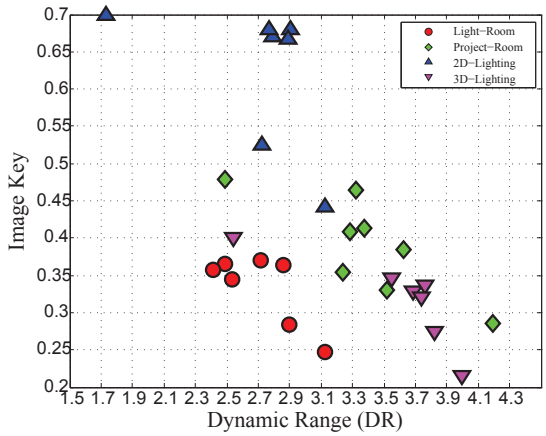

Figure 2: Dynamic Range Vs Image-Key plots for (a) Light-Room(LR). (b) Project-Room. (c) Lighting 2D. (d) Lighting 3D [5].

critical points in LDR feature detection, and are those for which HDR technology could bring most benefits.

Project Room (PR). The setup is composed of 8 different lighting scenes created by blocking light coming from a projector with the help of different objects. For each case, images with varying exposure time were captured using a Nikon D3100 digital camera. The setup is composed of several bright and dark colored objects arranged so as to create sharp shadows and overexposures in detailed areas. Created shadows hide the minute details for, e.g., bottom prints on memento, web-cam box printings etc.

Light room (LR). The dataset is composed of 7 different natural lighting conditions built by changes in global lighting due to opening and closing of window blinds, room ambient illumination and a diffused lighting from a tungsten lamp. For each condition, 6 images with different exposure time were shot using a Canon EOS 600D. This setup is also composed of dark and light objects with different type of object surfaces.

Both datasets are calibrated to the true physical luminance using the Minolta LS-100 Luminance meter, and can be downloaded from http://perso.telecom-paristech.fr/~gvalenzi/download.htm. We report in Figure 2 the variation in *dynamic range* of each scenes of our dataset. Image key [18] takes values on $[0, 1]$ and gives a measure of the overall brightness of the scene. Dynamic range is defined as $\log_{10}(L_{\max}/L_{\min})$, where $L_{\min}$ and $L_{\max}$ are the minimum and maximum HDR brightness values, respectively. Both properties give an indication of the variety of illumination conditions contained in the dataset.

B. Feature point detection

Feature extraction has been studied in vast details in computer vision literature where several techniques have been proposed and evaluated [2] taking into account different challenging transformations. In this paper, we focus on the two most widely used interest point detection schemes, i.e., *corner* and *blob* detectors, which are often used in several real time applications. In spite of several existing schemes for these approaches, we select two common detectors that have been used in similar evaluations for LDR content. For corner interest point detector, we employ the popular *Harris corner point* detector [3], 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* [4] detector.

Abbreviations	Description	L/G
D — Drago	An Adaptive logarithmic mapping [8]	G
W — Ward	Mapping based on histogram adjustment [19]	G
A — Ashikhmin	Gradient based mapping algorithm [20]	L
C — Chiu	Spatially non-uniform scaling algorithm [7]	L
M — Mantiuk	Perceptual method for contrast processing [21]	L
F — Fattal	Gradient domain HDR compression [9]	L
P — Pattnaik	Adaptive gain control for HDR [22]	L
R — Reinhard	Photographic tone reproduction method [23]	L
S — Schlick	Quantization techniques for visualization [24]	L

Table I: Local(L) and Global(G) tone mapping methods

C. Considered LDR/HDR modalities

For each illumination change dataset, we consider the following low and high dynamic range image modalities:

- **LDR** best exposed image: we take the subjectively best LDR exposure shot for each illumination setup, i.e., the one that a human surveillance operator would select based on large details with smallest area of over- or under-exposed pixels;
- Tone-mapped image: we consider two global **(GTM)** and seven local **(LTM)** TMO's (see Table I) to convert HDR pictures to 8-bit LDR, which are representative of the most popular tonemapping techniques for rendering HDR on LDR displays proposed in the literature;
- HDR linear values **(HDR-Lin)**, i.e., photometric luminance values stored in the HDR file;
- HDR perceptually encoded values: we consider a simple logarithmic **(HDR-Log)** encoding, according to Weber-Fechner's law; or the perceptually uniform encoding **(HDR-PU)** proposed in [25], which accounts for the drop of sensitivity at lower luminance levels. Notice that PU encoding needs photometrically calibrated HDR pixels as input. Both Log and PU values are rescaled in $[0, 1]$.

In total, 13 different image formats are thus considered for each lighting condition. We stress the difference between HDR encoded values and GTM pixel values: the former are the result of a simple transfer function and are not encoded on 8 bits; the latter, instead, are the result of a content-dependent operation, and are encoded on 8-bit, integer precision.

IV. Experimental Results and Discussion

Experiments are carried out on proposed datasets (LR,PR) and lighting dataset (2D,3D) of [5]. The only measure of accuracy considered is the repeatability rate (R-score) for each detector: Harris and SURF. In general, R-score is a standardized measure detailed in [2] to measure the detector

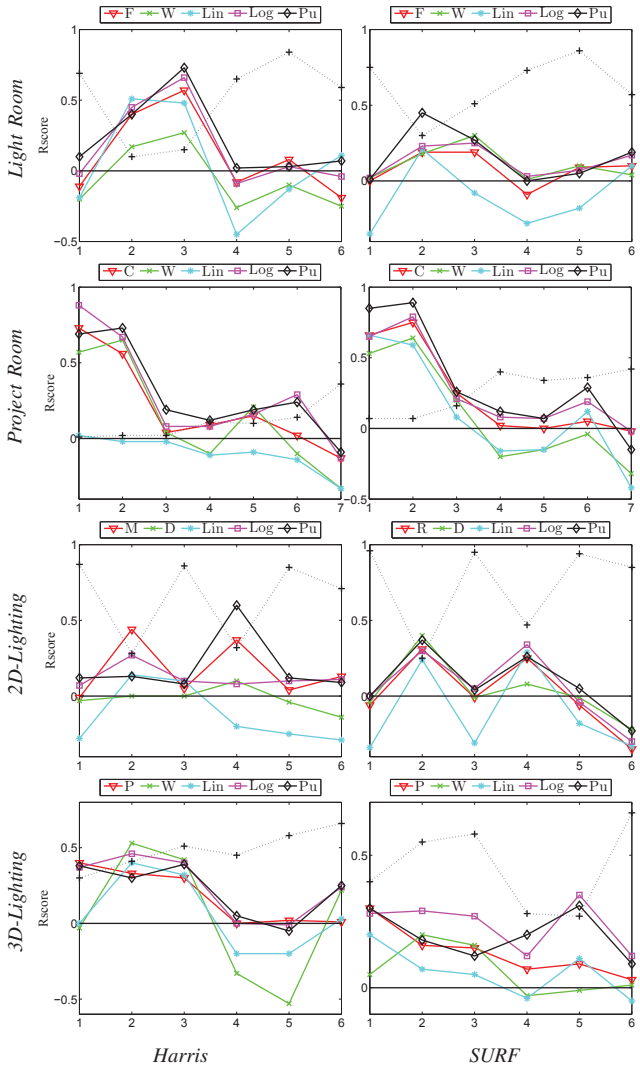

Figure 3: Relative gains by best LTM, GTM (abbreviated using Table I), Linear, Log and Pu HDR encodings with respect to LDR for different test datasets (scenes indicated by progressive numbers on x-axis). The dotted line shows the absolute R-scores of LDR.

accuracy, given as the fraction of feature points detected in the reference image which are repeated in the test 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. Higher R-score stands for better accuracy. For our evaluation, we used $\epsilon = 35$px, which is less than 1% the image size, similar to [5]. Also the evaluation scheme is confined to the strongest 200 key-points in marked ROI's (Region of Interests). This not only limits the feature point detection in pertinent areas, but also helps to ensure a fair comparison of the blob or corner key-point detection on diverse datasets, as different detectors result in highly different number of keypoints.

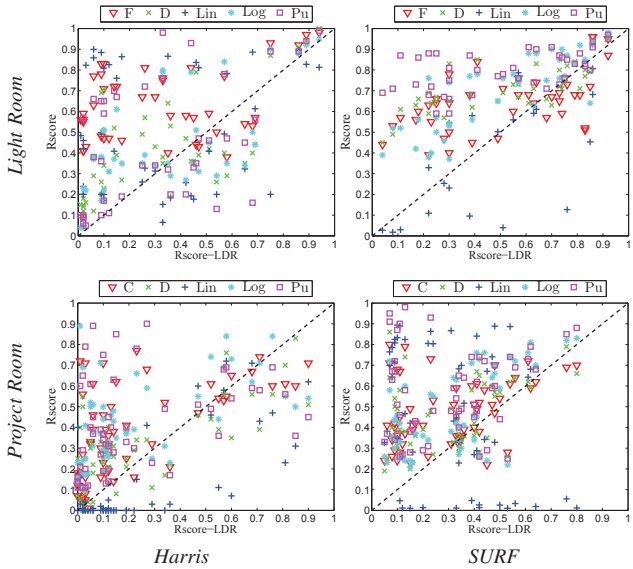

Figure 4: Scatter plots for HDR based formats (TMOs' abbreviated using Table I) with respect to LDR on proposed dataset using detector: Harris and SURF

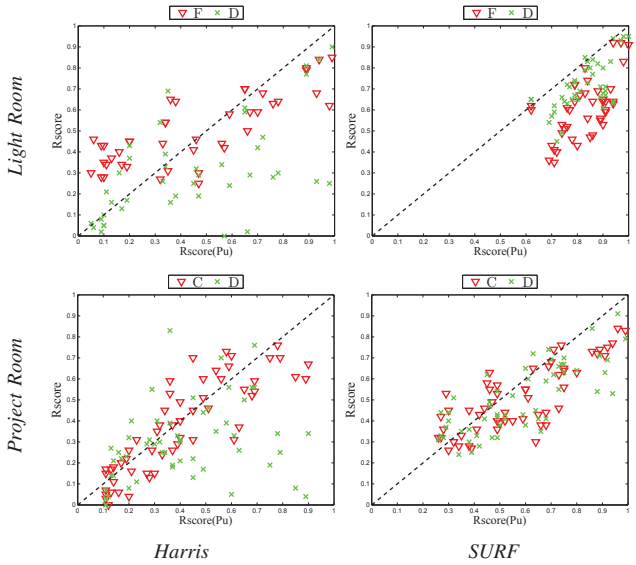

Figure 5: Scatter plots for best performing GTM and LTM (abbreviated using Table I) with respect to HDR-Pu encoding on proposed dataset using detector: Harris and SURF

The experimental study is conducted in two phases. In the first phase, for each dataset, one scene is selected as a reference image, and the repeatability is computed with the other scenes (test images). Relative gains are recorded for all best HDR based modalities (GTM, LTM, and HDR encoded formats) with respect to LDR, by subtracting the LDR R-score from each individual format as shown in Figure 3. The black dotted line depicts the absolute LDR R-score for each test image pair. For each dataset and using either detector, we observe high relative gains by HDR based modalities

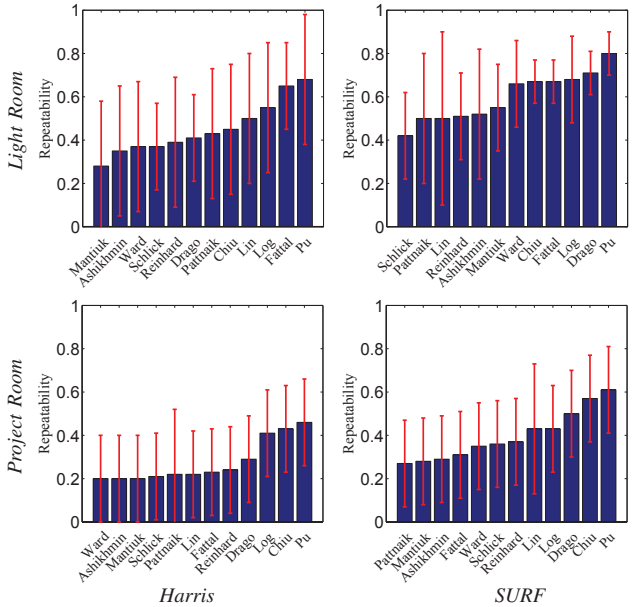

Figure 6: Average gain recorded by different formats over the LDR.

(especially by HDR-PU), but still, they are not positive everywhere (e.g., scene 0(ref) – 7 of Project Room dataset and scene 0(ref) – 6 of 2D-Lighting dataset test pairs from Figure 3).

In the second phase of the experiments, in order to determine more concrete quantitative information about all such possible cases, we expand our experimental test bench by involving all the possible images pairs for both LR and PR dataset, i.e., each condition is in turn the reference and the others are the test images. In this phase, we firstly determine the relative performance of the best performing HDR based formats, i.e., encoded HDR, LTMs and GTMs, with respect to the traditional LDR, producing the scatter plots shown in Figure 4. Each scatter plot shows the relative gain with respect to a compared format, while the dashed line is the 45° line: points lying above this line shows higher performance and points lying below are performing lower than the compared format. The distribution of points of tone mapped and encoded HDR based formats above the line, implies that these formats are capable to capture wider range of information from the images than the respective LDR format as shown in Figure 4. However, this is not true for HDR-linear format which shows the worst overall performance. On the other side, we also investigate the relative performance of the tone mapping and best HDR encoded format in Figure 5. The results obtained suggests that in many cases, applying a TMO entails a loss of detected keypoints.

In addition, the averages and standard deviations of the gains in R-score of all HDR modalities over LDR are shown in Figure 6. These are obtained by subtracting the R-score of the individual format from the absolute LDR R-score. In the following, we comment on the performance of the different HDR and LDR modalities for feature point detection.

HDR versus LDR. In all conditions and for both keypoint detectors, average values show significant gains of HDR or tone-mapped images over single LDR exposure. This is consistent with what has been found in [6]. However, based on the results from scatter plots in Figure 5, we observe that there are some scenarios where LDR records higher performance than the rest of the HDR based formats. We believe that this is mainly due to significant illumination differences in pertinent regions of image pairs.

HDR encodings. The best average repeatability scores are in general obtained with PU-HDR encoded values. This is not surprising, as HDR formats store most of the pertinent information in the scene, and it is therefore promising to research towards application of feature extraction on these modalities. From the results, it is also clear that these encodings give significantly better results than photometric HDR-Lin. This is a non-obvious conclusion of this work, i.e., that HDR-Lin is not appropriate to be used for feature extraction algorithms, especially by observing huge variation in its behavior. This seems to suggest that feature extraction algorithms designed and optimized for LDR content somehow require the perceptually scaled pixel values. Also for such algorithms, at least the parameters have to be completely retrained to cope with photometric luminance values.

HDR versus TMO's. Average R-scores gains over LDR by tone mapping techniques are either comparable or lower than those of HDR encoding formats. This partially accounts to information loss during their 8-bit quantization. However better performances recorded by some TMO's draws significant attention, such as Drago [8], Chiu [7] and Fattal [9]. In addition to the performance evaluation for LTMs', it is interesting that the gradient-based local techniques, i.e., Fattal and Chiu TMO's, have shown comparable gains than other LTM techniques, in specific scenarios. This is inverse to observations in perceptual applications [12], where these two LTMs' are deemed as worst performers. This further establishes that there is less congruency between visually pleasing tone mappings and vision-task-based optimal mapping technique. Another important point to note here is that these tone mappings perform better with blob detectors than corner point, and it is consistent with the observations in the literature comparing detectors [2]. In addition to all the observations, it is also worth mentioning that there is no unanimous winner amongst these tone mapping techniques using either detection criterion.

V. Conclusion

In this paper, we have presented a comprehensive evaluation of different HDR and LDR based modalities for visual feature detection, under changes in illumination conditions. The analysis of repeatability scores on different scenes

confirms the potential of HDR techniques over single LDR exposures. Furthermore, we have also observed that local TMO's producing very appealing results in terms of rendering quality are not necessarily the best option for image analysis. More interestingly, we have measured a consistent gain when using direct HDR pixel values over tonemapping and then extracting features, provided that HDR photometric pixels are first encoded to be approximatively linear to perception. This suggests that there might be quite a large room for improvement in feature extraction performance by designing optimal encoding schemes for HDR which can ensure high detection rates and that can be easily fused with current detection and recognition algorithms.

REFERENCES

[1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH*, 1997, pp. 369–378.

[2] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vision*, pp. 151–172, Jun. 2000.

[3] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proc. of Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[4] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision ECCV 2006*, 2006, pp. 404–417.

[5] B. Přibyl, A. Chalmers, and P. Zemčík, "Feature point detection under extreme lighting conditions," in *Proceedings of the 28th Spring Conference on Computer Graphics*. ACM, 2012, pp. 143–150.

[6] L. Chermak and N. Aouf, "Enhanced feature detection and matching under extreme illumination conditions with a hdr imaging sensor," in *IEEE 11th Int. Conf. on Cybernetic Intelligent Systems*, Aug 2012, pp. 64–69.

[7] K. Chiu, C. H. Shirley, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman, "Spatially nonuniform scaling functions for high contrast images," in *In Proceedings of Graphics Interface 93*, 1996, pp. 245–253.

[8] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," *Computer Graphics Forum*, pp. 419–426, 2003.

[9] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *ACM Trans. Graph.*, pp. 249–256, Jul. 2002.

[10] G. Kontogianni, E. K. Stathopoulou, A. Georgopoulos, and A. Doulamis, "HDR imaging for feature detection on detailed architectural scenes," *Int. Journ. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 325–330, 2015.

[11] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," *ACM Transactions on Graphics (TOG)*, pp. 640–648, 2005.

[12] M. Cadik, M. Wimmer, L. Neumann, and A. Artusi, "Evaluation of HDR tone mapping methods using essential perceptual attributes," *Computers and Graphics*, pp. 330 – 349, 2008.

[13] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, "Tone mapping-based high-dynamic-range image compression: study of optimization criterion and perceptual quality," *Optical Engineering*, pp. 102 008–102 008, 2013.

[14] A. Boschetti, N. Adami, R. Leonardi, and M. Okuda, "An optimal video-surveillance approach for HDR videos tone mapping," in *Signal Processing Conference, 2011 19th European*, Aug 2011, pp. 274–277.

[15] Y. Cui, A. Pagani, and D. Stricker, "Robust point matching in hdri through estimation of illumination distribution," in *Pattern Recognition*. Springer Berlin Heidelberg, 2011, pp. 226–235.

[16] L. Chermak, N. Aouf, and M. Richardson, "HDR imaging for feature tracking in challenging visibility scenes," *Kybernetes*, pp. 1129–1149, 2014.

[17] M. Rerabek, L. Yuan, L. Krasula, P. Korshunov, K. Fliegel, and T. Ebrahimi, "Evaluation of privacy in high dynamic range video sequences," pp. 92 170E–92 170E–13, 2014.

[18] A. O. Akyüz and E. Reinhard, "Color appearance in high-dynamic-range imaging," *J. Electronic Imaging*, p. 033001, 2006.

[19] G. W. Larson, H. Rushmeier, and C. Piatko, "A visibility matching tone reproduction operator for high dynamic range scenes," *IEEE Transactions on Visualization and Computer Graphics*, pp. 291–306, Oct. 1997.

[20] M. Ashikhmin, "A tone mapping algorithm for high contrast images," pp. 145–156, 2002.

[21] R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM Trans. Appl. Percept.*, pp. 286–308, Jul. 2006.

[22] S. Pattanaik and H. Yee, "Adaptive gain control for high dynamic range image display," in *Proceedings of the 18th Spring Conference on Computer Graphics*, ser. SCCG '02. ACM, 2002, pp. 83–87.

[23] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, pp. 267–276, Jul. 2002.

[24] C. Schlick, "An adaptive sampling technique for multidimensional integration by ray-tracing," in *Photorealistic Rendering in Computer Graphics*. Springer, 1994, pp. 21–29.

[25] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full dynamic range images," in *Human Vision and Electronic Imaging XIII*, ser. Proceedings of SPIE, January 2008, pp. 6806–10.