

Evaluation of Feature Point Detection in High Dynamic Range Imagery

Bronislav Příbyl^{a,*}, Alan Chalmers^b, Pavel Zemčík^a, Lucy Hooberman^b, Martin Čadík^a

^a*Department of Computer Graphics and Multimedia, Faculty of Information Technology,
Brno University of Technology, Božetěchova 2, 612 66 Brno, Czech Republic*

^b*WMG, University of Warwick, CV4 7AL Coventry, United Kingdom*

Abstract

This paper evaluates the suitability of High Dynamic Range (HDR) imaging techniques for Feature Point (FP) detection under demanding lighting conditions. The FPs are evaluated in HDR, tone mapped HDR, and traditional Low Dynamic Range (LDR) images. Eleven global and local tone mapping operators are evaluated and six widely used FP detectors are used in the experiments (Harris, Shi-Tomasi, DoG, Fast Hessian, FAST, and BRISK). The distribution and repeatability rate of FPs are studied under changes of camera viewpoint, camera distance, and scene lighting. The results of the experiments show that current FP detectors cannot cope with HDR images well. The best contemporary solution is thus tone mapping of HDR images using a local tone mapper as a pre-processing step.

Keywords: Feature point detection, Interest point detection, Corner point detection, High dynamic range imagery, Repeatability rate, Evaluation of tone mapping operators

1. Introduction

Many computer vision tasks, such as image analysis, registration and indexing, object tracking, 3D reconstruction, visual navigation (SLAM), etc. rely on the presence of low-level features in images [1]. These features typically are blobs, edges, or points. In the case of points, these include corner points, interest points, or most often *Feature Points* (FPs). These image points usually correspond to some real points in the scene although some of them might correspond to “deceiving phenomena”, such as reflections or shadow edges as well.

The detection of FPs is strongly dependent on the illumination of the scene at the moment of image capture [2]. Demanding lighting conditions or wrong camera settings can cause FP detectors to fail to detect many of the points. This is particularly true when dealing with images of the natural world where the average luminance levels may vary approximately between 10^{-3} cd/m² (on a starlit night) and 10^6 cd/m² (on a sunny day), see [3]. Such a difference between the luminance levels can generate a dynamic range of $1 : 10^9$, or 30 stops¹. When capturing images under demanding lighting conditions, one has to carefully set the camera and arrange the scene which is a limiting factor and sometimes cannot be performed completely successfully. An alternative approach is to use High Dynamic Range (HDR) imagery – a technology which has penetrated a significant segment of professional cameras in recent years and is now starting to appear even in low end consumer cameras and smart phones.

HDR imagery allows the capture and storage of greater dynamic range of light in a scene than traditional Low Dynamic Range (LDR) imagery. LDR imagery uses 8-bit integers to store pixel values, thus limiting the intensity range to 0 – 255 and the dynamic range to 8 stops. HDR

*Corresponding author at DCGM FIT BUT, Božetěchova 2, 612 66 Brno, Czech Republic. Phone number +420 777 241 447.

Email addresses: ipribyl@fit.vutbr.cz (Bronislav Příbyl), alan.chalmers@warwick.ac.uk (Alan Chalmers), zemcik@fit.vutbr.cz (Pavel Zemčík), l.hooberman@warwick.ac.uk (Lucy Hooberman), cadik@fit.vutbr.cz (Martin Čadík)

¹The term *stop* originates in the photographic community and denotes the exposure range of a scene in a power of 2 units. A dynamic range $1 : 2^n$ equals therefore a dynamic range of n stops. The terms *step* and *f-stop* are also used.

imagery, on the other hand, typically uses more than 8 bits, allowing a dynamic range up to hundreds of stops [3]. This is a fundamental advantage which allows HDR imagery to represent high ranges of lighting, providing far more detailed information about the scene. HDR thus has the potential to improve the performance of many computer vision tasks, including feature point detection.

Many evaluations of feature point detectors have been performed previously, both general ones [1, 2, 4, 5, 6] and application specific ones [7, 8, 9, 10]. To the best of our knowledge, all these evaluations have been carried out using classical LDR images only. Only a few recent papers have considered HDR imagery in FP detection, e.g. [11, 12], but no thorough comparison with LDR has been done so far. We, therefore, intend to answer the question ‘‘Can the use of HDR imagery be significantly beneficial for feature point detection and if so, why?’’

The rest of this paper is organized as follows: In Section 2, we describe the six FP detectors selected for our experiments; discuss the literature on previous comparisons of FP detectors; and review methods for tone mapping of HDR images. Section 3 details the setup for the evaluation. The results are presented and analysed in Section 4. Finally, we conclude and make suggestions for future work in Section 5.

2. Related Work

2.1. Feature Point Detectors

A number of FP detectors have been proposed in the literature. For a comprehensive survey, we refer the reader to [13]. The FP detectors are often used together with FP descriptors, e.g. SIFT [14], SURF [15], BRISK [16], which are beyond the scope of this paper. The following six widely used detectors are good representatives of the different approaches to FP detection.

Harris Corner Detector: This method is based on the local auto-correlation function reflecting local intensity changes in the image [17]. For each point \mathbf{x} , the second moment matrix

$$\mathbf{M}(\mathbf{x}) = \begin{bmatrix} I_x^2(\mathbf{x}) & I_x I_y(\mathbf{x}) \\ I_x I_y(\mathbf{x}) & I_y^2(\mathbf{x}) \end{bmatrix} \quad (1)$$

is computed, where I_x and I_y are the derivatives of intensity in the x and y directions at point \mathbf{x} . The components of the matrix \mathbf{M} are usually smoothed using a Gaussian to make the detection more robust. Then the point score $R(\mathbf{x})$ is computed as

$$R(\mathbf{x}) = \lambda_1 \lambda_2 - k \cdot (\lambda_1 + \lambda_2)^2 \quad (2)$$

where λ_1 and λ_2 are the eigenvalues of $\mathbf{M}(\mathbf{x})$ and k is a sensitivity factor. Since direct computation of the eigenvalues is expensive, Harris and Stephens introduce an approximation of Eq. (2) by means of the determinant and the trace of $\mathbf{M}(\mathbf{x})$:

$$R(\mathbf{x}) = \det(\mathbf{M}(\mathbf{x})) - k \cdot \text{tr}(\mathbf{M}(\mathbf{x}))^2 . \quad (3)$$

Shi-Tomasi: The minimum eigenvalue detection method proposed by Shi and Tomasi [18] relies on the same second moment matrix \mathbf{M} as the Harris detector does, but explicitly computes its eigenvalues according to Eq. (2) unlike Harris. This results in higher computational demands but also in feature points which are more stable for tracking.

DoG: The Difference of Gaussian is the detector part of the so called SIFT (Scale Invariant Feature Transform) combined feature detector and descriptor proposed by Lowe [14]. In this paper, we only use the detection part. This detector is multiscale, which is achieved by building a scale space

$$\mathcal{L}(\mathbf{x}, \sigma) = g(\sigma) * I(\mathbf{x}) \quad (4)$$

at each point \mathbf{x} and scale σ as the convolution of the Gaussian $g(\sigma)$ with an image I . Feature points are detected as extrema in the difference of Gaussian function $D(\cdot)$ convolved with the image, which can be computed from the difference of two nearby scales separated by a constant factor k

$$D(\mathbf{x}, \sigma) = \mathcal{L}(\mathbf{x}, k\sigma) - \mathcal{L}(\mathbf{x}, \sigma) . \quad (5)$$

Fast Hessian: This is the detector part of the so called SURF (Speeded up Robust Features) combined feature detector and descriptor proposed by Bay et al. [15]. In this paper, we only use the detection part. This detector approximates the Hessian matrix

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (6)$$

at each image point \mathbf{x} at scale σ . $L_{xx}(\mathbf{x}, \sigma)$ is the convolution of the Gaussian second order partial derivative $\frac{\partial^2}{\partial x^2}g(\sigma)$ with image at point \mathbf{x} and similarly for $L_{xy}(\mathbf{x}, \sigma)$ and $L_{yy}(\mathbf{x}, \sigma)$. A scale space is thus created by applying filters with increasing σ . A $3 \times 3 \times 3$ -neighbourhood non-maximum suppression [19] is then applied in the scale space to filter the strongest feature points.

FAST: The Features from Accelerated Segment Test (also called “local intensity comparison”) method by Rosten and Drummond [20] considers a pixel to be a possible corner point if it has n contiguous surrounding pixels on a circle, which are either brighter or darker than the central pixel. The value of n effectively controls a threshold angle θ which describes which features will be detected (both corners and edges or just corners). The circle considered usually has a radius of 3 pixels in practical applications.

BRISK: The Binary Robust Invariant Scalable Keypoints method by Leutenegger et al. [16] is a combined feature detector and descriptor. In this paper, we only use the detection part. It is a multiscale detector, utilizing the FAST detector at each scale. The scale of each feature point is obtained in the continuous domain via quadratic function fitting.

The fundamental part of any FP detector is the computation of some kind of derivative of the pixel values encoded in an image. The bigger the magnitude of the derivative, the stronger the feature point detected. The design of existing FP detectors (e.g., the derivative thresholds) assumes a display-referred LDR image, usually gamma-corrected. In this case, the magnitudes of the derivatives of pixel values in dark and bright regions of the image would not be significantly different. However, in a scene-referred HDR image the pixel values encode linear luminance of the real scene². The derivatives in the HDR image therefore increase significantly in the bright areas and can be orders of magnitude bigger than in the dark areas. Imagine, for example, a step edge printed on a piece of paper. While the reflectivity (and derivative) of the edge remains constant in the real world, the reflected light (captured in an HDR image) is proportional to the reflectivity *and illumination* of the patch. The strongest feature points would thus be detected primarily in highly illuminated areas leaving the FPs in dark areas undetected, which is generally useless. We assume that all the contemporary FP detectors would process HDR images inefficiently due to this fact³. Our measurements confirm this hypothesis, see Section 4.

2.2. Comparison of FP Detectors

A number of papers on the comparison and evaluation of feature point detectors have been published in the last decade. Schmid et al. [1] presented an extensive study, where they compared FP detectors on two planar scenes under changes in rotation, viewpoint and illumination, and artificially added image noise. They were the first to introduce and evaluate the *Repeatability Rate* (abbreviated as RR) which describes (in-)dependence of FP detection on imaging conditions. The feature is “useful” if it is detected in one image and it is visible in the other image

²The pixel values are typically proportional to luminance (with an unknown scaling factor) for uncalibrated HDR images, or they represent scene luminance in candelas per square metre (cd/m²) in calibrated images.

³Humans, on the other hand, are able to cope with the vast range of luminance values by means of *visual adaptation* mechanisms. Visual systems need to adapt to the background illumination to be able to distinguish objects. This behavior is measured in detection threshold experiments, where the difference dL between stimulus and background luminance increases in proportion to the background luminance L resulting in non-linear threshold-versus-intensity (TVI) function. The *linear part* of TVI is known as Weber’s law [21], which in this case states that the contrast sensitivity is constant ($dL/L = \text{const.}$). This implies that human response to luminance may be roughly approximated by a logarithmic function ($\ln L$). However, due to complexity of human visual system, other compressive nonlinearities (e.g., a power function) may be more appropriate depending on observation conditions.

as well. The feature is “repeated” when it is detected close to the same real-world position in the other image. The RR of features in an image pair is denoted r and it is defined as the ratio between the number of repeated features n_{repeated} and the number of useful features n_{useful} :

$$r = \frac{n_{\text{repeated}}}{n_{\text{useful}}}. \quad (7)$$

Mikolajczyk et al. [2] compared affine-invariant *region detectors*. The scenes used for their experiments were also planar or near-planar. Fraundorfer and Bischof [4] extended the work of Mikolajczyk et al. by introducing a new tracking method allowing the detectors to be evaluated on *non-planar scenes* as well. They discovered that non-planar scenes exhibit significantly lower repeatability rate of FP detectors than planar scenes. Rodehorst and Koschan [5] compared performance of detectors on a set of planar and 3D scenes. The authors addressed the problem of non-uniform distribution of FPs in the image and proposed a solution by using an *adaptive detector threshold*. Moreels and Perona [6] explored the performance of combinations of detectors and descriptors with a test bed of 3D objects. They exploited *geometric constraints* between triplets of views.

More recently, several application-specific studies of FP detectors have been published. The studies naturally focus on specific parameters of the detectors. For example, Gil et al. [8] evaluated behavior of the detectors and descriptors with respect to their use in visual navigation. Accordingly, they studied the stability of detected features throughout the whole *image sequences* rather than just in image pairs. Rosten et al. [22] redefined the RR to express the stability of features throughout an image sequence. The RR of features in an image sequence is denoted R and it is defined as:

$$R = \frac{N_{\text{repeated}}}{N_{\text{useful}}}, \quad (8)$$

where N_{repeated} and N_{useful} are numbers of features summed over all image pairs in an image sequence:

$$N_{\text{repeated}} = \sum_{i \in \text{seq}} \sum_{j \in \text{seq}; j \neq i} n_{\text{repeated}}, \quad (9a)$$

$$N_{\text{useful}} = \sum_{i \in \text{seq}} \sum_{j \in \text{seq}; j \neq i} n_{\text{useful}}, \quad (9b)$$

where i and j denote images from the sequence seq .

Jazayeri and Fraser [9] concentrated on feature-based matching and close range photogrammetry. They were also amongst the first to evaluate the *speed* of the detectors. Gauglitz et al. [10] compared the detectors and descriptors with regard to real-time visual tracking; concentrating on *execution time* and *precision*.

Recently, Kontogianni et al. [12] published a brief study comparing the number of FPs detected by the DoG, Fast Hessian, FAST, and ORB detectors in images of architectural scenes. The scenes were captured into single exposure LDR images and into HDR images tone mapped using the method of Mantiuk et al. [23]. They came to the conclusion that the tone mapped images yield more FPs than LDR images, which is in agreement with our findings in the 3D scene, see Section 4.1. In addition, May et al. described a study of object recognition using HDR [24] and image fusion [25]. However, they evaluated only the number of FPs in the same views and not the *repeatability and stability* that heavily affect the FP applicability.

To the best of our knowledge, no study of the influence of the HDR and tone mapping on the *repeatability and stability of FP detectors* has been presented so far.

2.3. Tone Mapping of HDR Images

Most of the implementations of contemporary computer vision algorithms work with ordinary (LDR) images and they are not designed for HDR images, as was pointed out in Section 2.1. Two possible solutions to this problem exist: (1) Adaptation of existing algorithms to be able to work with HDR data, which is a promising avenue for future research; or (2) preceding current methods with a suitable conversion from the HDR to LDR image, known as tone mapping.

Methods performing tone mapping (tone mapping operators, TMOs) are typically classified into *global* tone reproduction curves and *local* tone mapping methods. Definitions of these two classes are not fully consistent. For our purpose, we adopt the most commonly used definition, see [3, 26, 27]:

Global TMO: The mapping function is fixed for all pixels in an image. The same input pixel intensities are thus mapped to the same values.

Local TMO: The mapping function can vary depending on a local neighbourhood of the mapped pixel. The same input pixel intensities may be mapped to different values.

The classification into global and local TMOs is suitable also from the FP detection point of view: Global and local operators may change the local contrast differently. Since the detection of FPs is often based on derivation magnitude, which depends on local contrast, images tone mapped with global and local operators are expected to exhibit different behaviors.

Another dimension worth exploring is the visual coherency of tone mapped images. Many TMOs can automatically tune their parameters, but they do it independently for each image. This potentially results in different image appearance in the set of images under consideration, which may in turn decrease the performance of FP detectors. We have thus included a temporal TMO into our evaluation, treating the images as if they were part of a video sequence.

We will now briefly describe the representative TMOs we used in this paper to assess the effects of tone mapping on FP detection. The names in round brackets are the abbreviations used throughout the rest of the paper.

Logarithmic Mapping (gLogMap) is based on the application of a *logarithmic function* to the input luminance values. The most general form of logarithmic mapping is

$$L_o(\mathbf{x}) = \frac{\log(1 + qL_i(\mathbf{x}))}{\log(1 + kL_{i,\max})}, \quad (10)$$

where L_i and L_o are the input and output luminance, respectively, and k and q are constants for tuning the desired appearance of the output image.

Histogram Adjustment (gLarson97) by Ward Larson et al. [28] uses a *modified histogram equalization* technique [29, ch. 3]. A \log_2 -histogram of an image is calculated and subsequently equalized. Finally, advanced aspects of human vision are simulated, such as glare, color sensitivity or visual acuity.

Photographic Tone Reproduction - global part (gReinhard02) proposed by Reinhard et al. [30] is motivated by the time-tested techniques of photographic practice of dodging-and-burning and the zone system. The method consists of *purely global component* (gReinhard02) and an optional local part, which we analyze separately (lReinhard02). The global component compresses mainly high luminance values as follows: $L_o(\mathbf{x}) = \frac{L_m(\mathbf{x})}{1 + L_m(\mathbf{x})}$, where L_m is the input luminance scaled by the scene key value (geometric mean).

Dynamic Range Reduction Inspired by Photoreceptor Physiology (gReinhard05) is a global TMO [31] that aims to mimic visual adaptation process which occurs in photoreceptors on human retina. The input color channels are individually compressed using straightforward *sigmoidal function* with perceptually justified parameters.

Display Adaptive Tone Mapping (gMantiuk08) automatically adjusts the HDR conversion given the characteristics of the display device [32]. To that end, human visual system models of scene and display observers are computed and used to steer a global *piece-wise linear tone reproduction curve*.

Photographic Tone Reproduction - global + local part (lReinhard02) represents the complete pipeline of the Photographic TMO [30], i.e. (gReinhard02) subsequently locally enhanced with the automatic dodging-and-burning. The popularity of this TMO is due to fairly stable automatic parameter estimation which makes the method easy to use in practice.

Gradient Domain HDR Compression (lFattal02) *modifies the gradient field* of the HDR image to achieve dynamic range compression [33]. More specifically, large gradients are attenuated, while the small ones (fine details) are preserved resulting in detail-rich and crisp outputs. The attenuation function is blended from multiple scales to prevent halo-like artifacts. The final image is obtained from the modified gradient field using Poisson’s equation solver.

Contrast Mapping (lMantiuk06map) is based on *multiscale processing of HDR image contrast* [23]. Specifically, the contrast is represented using a Gaussian pyramid in the logarithmic space. This way, not only local, but also global contrasts (low spatial frequencies) may be modified. Perception-motivated monotonic *transducer functions* are proposed to simulate human responses to physical contrasts.

Contrast Equalization (lMantiuk06equ) performs *equalization of histogram of contrasts* to produce highly detailed, sharp and crisp images [23]. The contrasts are represented in a way similar to (lMantiuk06map). After the equalization step, the modified contrasts are converted back to luminances by solving an optimization problem over all scales.

Edge-Avoiding Wavelets (lFattal09) involves a *wavelet lifting scheme* to build an edge-preserving multiscale decomposition of an HDR image [34]. The adaptive wavelet functions make fast construction of such an image representation possible which in turn enables band-independent multiscale processing in real time. The range compression is achieved by modifying the wavelet (detail) coefficients using the following function: $f(x) = \text{sign}(x)|x|^\beta$, where x is a detail coefficient and $\beta < 1$.

Real-Time Automated TM System for HDR video (gKiser12(t)) is a global *temporal TMO* based on (gReinhard02) [35]. To stabilize the TMO parameters of the current frame the method exploits information from previous frames in the HDR video sequence. Specifically, a “leaky integrator” [36] is applied on parameters of global tone reproduction curve (gReinhard02) which efficiently eliminates flicker from tone mapped videos.

The selected methods represent major trends in tone mapping research: global tone reproduction curves (gLogMap, gLarson97, gReinhard02, gReinhard05, gMantiuk08, gKiser12(t)), local tone mapping operators (lReinhard02, lFattal02, lMantiuk06map, lMantiuk06equ, lFattal09), multiscale optimization frameworks (lMantiuk06map, lMantiuk06equ, lFattal09), gradient-based methods (lFattal02), perceptually-motivated approaches (gLarson97, gReinhard05, lMantiuk06map, gMantiuk08), and temporal tone mapping operators (gKiser12(t)). Please refer to HDR textbooks, such as [3, 37] for more in-depth information.

3. Experimental Setup

In this section, a detailed description of our experiments is given. We first describe the scenes created for the evaluation, how they were captured, and into which image formats were they stored. Next, methods of feature point detection are described and the evaluation methodology is defined.

3.1. Scenes

Planar and non-planar scenes cause different behavior of FP detectors, as shown by Fraundorfer and Bischof in [4]. Two test scenes have therefore been created, see Fig. 1.

- A planar (2D) scene containing three different paper posters next to each other, attached to the same plane.
- A 3D scene containing several non-planar rigid objects.

These scenes were placed into a totally dark room and illuminated by two 2kW Arri tungsten lights to create extreme lighting conditions. Fig. 2 shows the light configurations. The 2D scene comprised a dark poster placed into a shadow, another poster lit by one light and a bright poster lit by both lights. The 3D scene was made analogously to the 2D scene: A coarse dark statuette

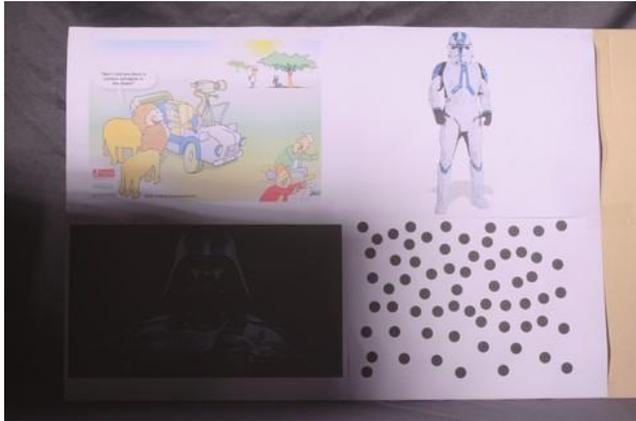


Figure 1: The test scenes used in the experiments: a 2D scene (containing just one plane; left) and a 3D scene (containing non-planar objects of different shapes; right)

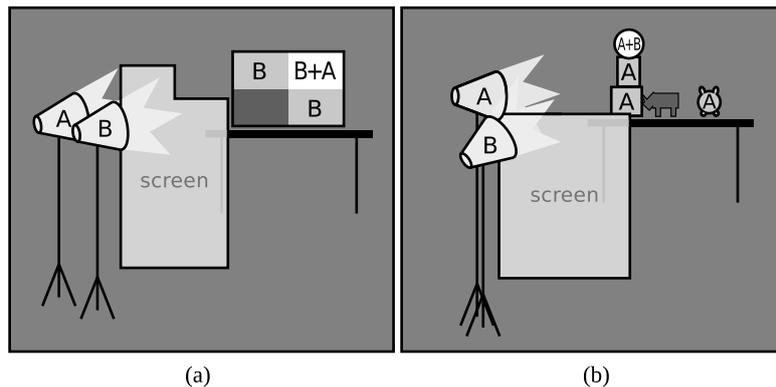


Figure 2: Configuration of the two lights, A and B, relatively to the 2D scene (a) and to the 3D scene (b). Individual parts of the scenes are labeled with “A”, “B” or “A+B” according to the sources of illumination of each part

of a rhino placed into a shadow, a puppet lit by one light, and a ball of creased white paper lit by both lights. Both scenes have been designed to generate as much dynamic range of light as possible while containing textured areas suitable for detection of FPs. The 2D scene has an average dynamic range of 12.1 stops and the 3D scene, 13.7 stops, as can be seen in Fig. 3. Both scenes also contain a dot pattern printed on paper used for automatic calibration of internal and external parameters of the camera.

3.2. Image Sequences

The scenes have been captured in three different image sequences by changing the camera viewpoint, distance, and scene lighting. Example images from the sequences can be seen in Fig. 4 and Fig. 5.

Viewpoint changing sequence. The camera moved along a circle with a center point in the middle of the scene with a 2.5° step. Since the scenes were captured 21 times, the total angular range of viewpoints was 50° .

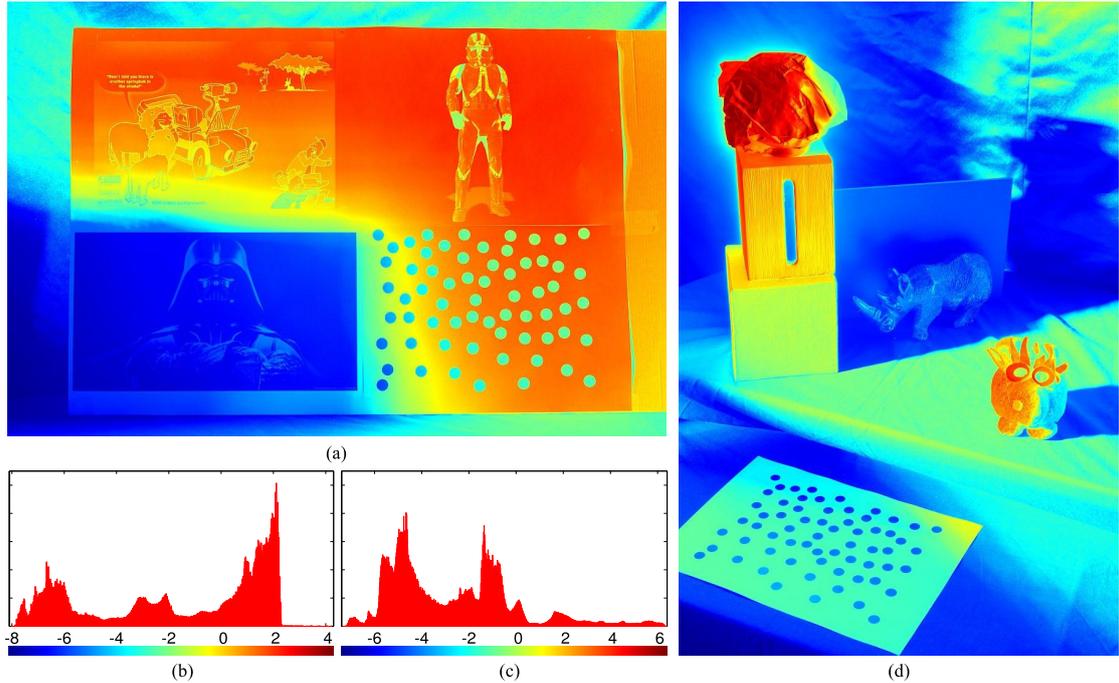


Figure 3: Luminance distributions and \log_2 -luminance histograms of example HDR images used in the experiments. The images are shown in false colors to depict luminance levels present in the scenes (dark blue = lowest luminance, dark red = highest luminance). (a) HDR image of the 2D scene in false colors, (b) \log_2 -luminance histogram of the 2D scene image, (c) \log_2 -luminance histogram of the 3D scene image, (d) HDR image of the 3D scene in false colors

Distance changing sequence. The scene was captured 7 times and the distance between the camera and the scene increased exponentially, yielding the distance sequence of 100, 103, 109, 122, 147, 197 and 297 cm. The distance was chosen to increase exponentially rather than linearly to examine the behavior of FP detectors under subtle as well as dramatic distance changes.

Lighting changing sequence. The scenes were also captured 7 times, each time with different combination of 3 light sources being on or off, with at least one of them on. The light sources used were the strip lights in the room and the two 2kW tungsten lights mentioned above. When referencing lighting conditions in the experiments a three-digit binary code is used to represent it where each digit represents the state of one light source in the order: (1) strip lights, (2) tungsten light A, (3) tungsten light B. For example, the code *101* means that the strip lights and tungsten light B were on. For the configuration of lights see Fig. 2.

3.3. Image Formats

In this section, we describe the formats used to store the image data. We also describe how the individual formats have been created, which is illustrated in Fig. 6.

To capture all the dynamic range in the scenes, several JPG images were captured at different exposure levels at each camera position. A Canon EOS-1Ds Mark II camera mounted on a tripod was used for this purpose. All the images have a resolution of approx. 16 Mpx.

The multiple exposures were then combined into a single radiance map using the method of Debevec and Malik [38]. An approximate photometric calibration was performed using the multiplication factor obtained by the fraction between the measured illumination in the scene and the recovered value in the radiance map. This calibrated radiance map was then tone mapped as described later. However, in practice the HDR images are rarely calibrated and the range of values is virtually arbitrary. To reflect this fact and to allow a fair comparison with LDR formats stored in JPG, we added a linearly re-scaled HDR image ($[0.0 - 256.0]$) into the test set. This scaled HDR image and the calibrated radiance map were stored using the Radiance (.hdr) format [37, p. 95].

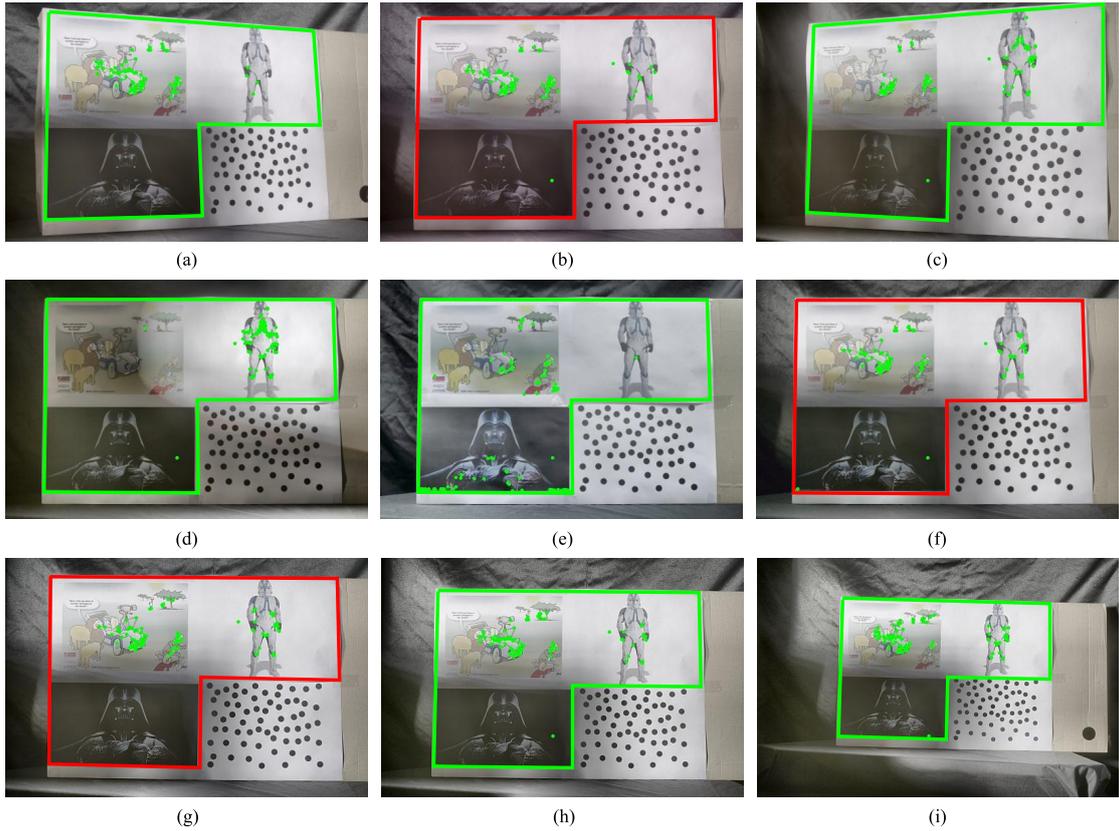


Figure 4: Example images from sequences of the 2D scene. Viewpoint changing sequence: -25° (a), 0° (b) and $+25^\circ$ (c). Lighting changing sequence: lighting conditions 010 (d), 100 (e) and 111 (f). Distance changing sequence: 100 cm (g), 109 cm (h) and 147 cm (i). Feature points detected using the Harris corner detector are marked with green dots. The feature points were detected only inside the Regions Of Interest (ROI) which are marked as green and red polygons. Images with red ROIs are the reference images for each sequence (b, f, g). The dot pattern is used solely for the purpose of automatic camera calibration

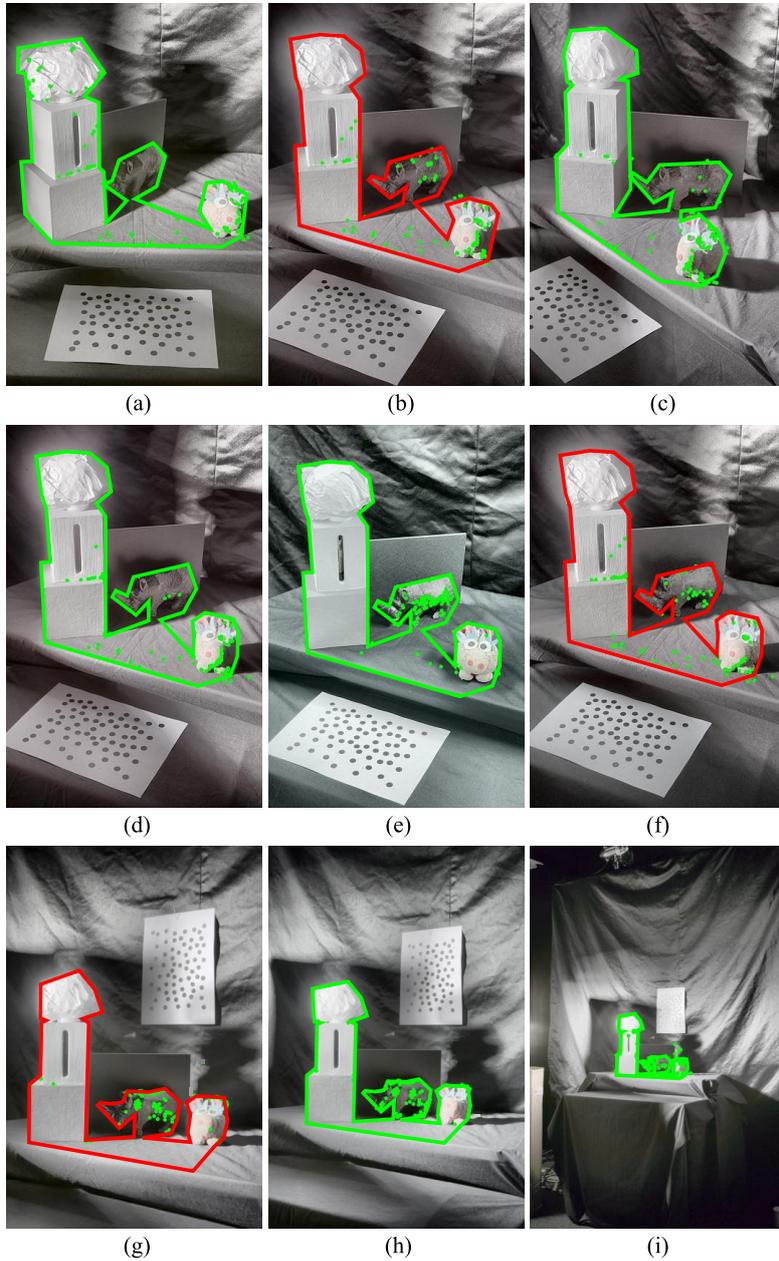


Figure 5: Example images from sequences of the 3D scene. Viewpoint changing sequence: -25° (a), 0° (b) and $+25^\circ$ (c). Lighting changing sequence: lighting conditions 010 (d), 100 (e) and 111 (f). Distance changing sequence: 100 cm (g), 122 cm (h) and 297 cm (i). Feature points were detected using the Harris corner detector are marked with green dots. The feature points were detected only inside of the area approximately marked as green and red polygons. Images with red polygons are the reference images for each sequence (b, f, g). The dot pattern is used solely for the purpose of automatic camera calibration

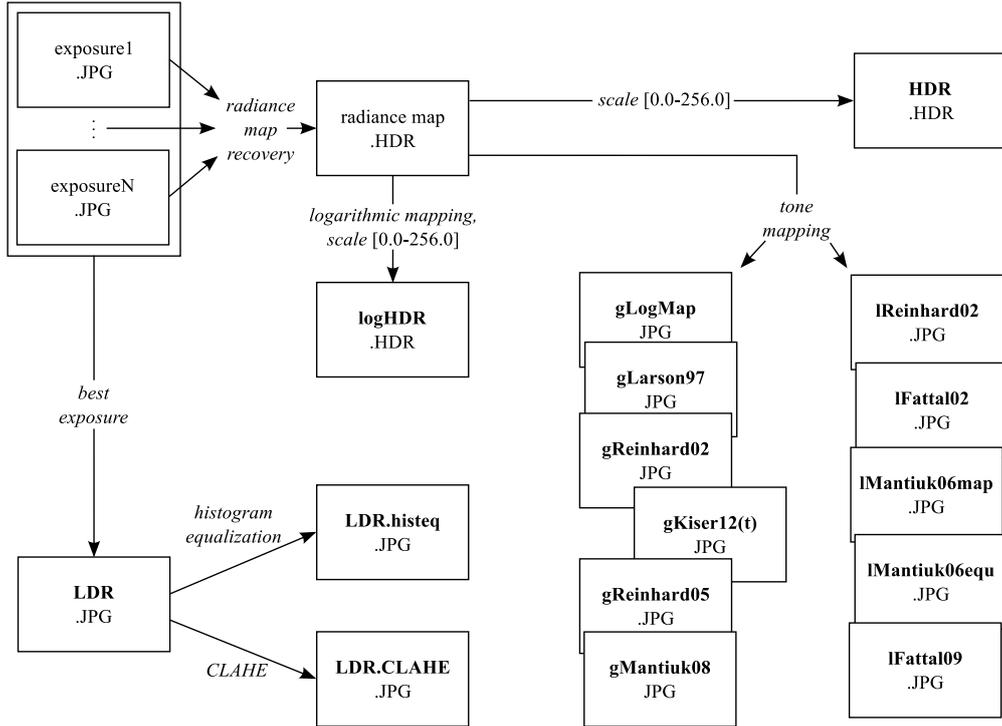


Figure 6: Workflow of image formats generation. Each format is depicted by a box with name and file extension used. Bold formats were used in the evaluation (non-bold ones are the intermediate formats). The names of the tone mapped formats consist of a prefix (**g** for global, **l** for local tone mapper), first author name and year (e.g. **Mantiuk06**) and an optional postfix (e.g. **(t)** for a temporally coherent version)

We assumed that contemporary FP detectors would detect FPs only in highly illuminated areas of the HDR image due to linear encoding of scene luminance, see discussion in Section 2.1. Therefore, we transformed the luminance scale of the HDR image using several approaches. The first approach was to encode the HDR logarithmically, motivated by the Weber’s law [21]. This was achieved by applying logarithmic mapping to the calibrated radiance map, followed by linear scaling into the interval $[0.0 - 256.0]$. The other approach was tone mapping. Since it was not obvious which TMO produces the most suitable images for FP detection, we chose to apply five global and five local operators.

To analyze whether HDR and tone mapped imagery brings any benefits over “classical” LDR imagery, we also included single exposure images into the experiments. The best exposure was selected from the set of all exposures by maximizing the number of well-exposed pixels inside the ROIs.

Some of previous studies suggest that a pre-processing step should be used to enhance local contrast when detecting features in an LDR image as low contrast can limit performance of FP detectors. Jazayeri and Fraser [9] proposed to use Wallis filtering [39]. We have tested pre-processing of LDR images in our previous evaluation [40] and concluded Wallis filter is beneficial only for images with poor contrast which is not the case of images captured under extreme lighting conditions. We thus applied two more efficient contrast enhancing methods in this evaluation: histogram equalization and Contrast Limited Adaptive Histogram Equalization (CLAHE) [41].

All this resulted in scenes captured in sixteen different image formats at each camera position, see Fig. 6. The image formats are the following (here, the “HDR image” refers to the calibrated radiance map):

- **HDR:** HDR image linearly scaled into the interval $[0.0 - 256.0]$.
- **logHDR:** HDR image transformed using logarithmic mapping and scaled into the interval $[0.0 - 256.0]$.

- **LDR:** The best exposed “single exposure” directly from the camera.
- **LDR.histeq:** The LDR image after histogram equalization.
- **LDR.CLAHE:** The LDR image with histogram equalized using CLAHE [41].
- **gLogMap:** HDR image tone mapped using logarithmic mapping according to Eq. (10) with tuning parameters $k = q = 1$, gamma corrected.
- **gLarson97:** HDR image tone mapped using the Histogram Adjustment method of Ward Larsson et al. [28], gamma corrected.
- **gReinhard02:** HDR image tone mapped using the global part of Photographic Tone Reproduction method of Reinhard et al. [30], gamma corrected.
- **gKiser12(t):** HDR image tone mapped using the *temporal* TMO of Kiser et al. [35], gamma corrected.
- **gReinhard05:** HDR image tone mapped using the Dynamic Range Reduction Inspired by Photoreceptor Physiology method of Reinhard and Devlin [31], gamma corrected.
- **gMantiuk08:** HDR image tone mapped using Display Adaptive Tone Mapping of Mantiuk et al. [32].
- **lReinhard02:** HDR image tone mapped using the complete Photographic Tone Reproduction method of Reinhard et al. [30] (both global and local part), gamma corrected.
- **lFattal02:** HDR image tone mapped using the Gradient Domain HDR Compression method of Fattal et al. [33].
- **lMantiuk06map:** HDR image tone mapped using the Contrast Mapping method of Mantiuk et al. [23], gamma corrected.
- **lMantiuk06equ:** HDR image tone mapped using the Contrast Equalization method of Mantiuk et al. [23], gamma corrected.
- **lFattal09:** HDR image tone mapped using the Edge-Avoiding Wavelets method of Fattal [34], gamma corrected.

Images in the *HDR* and *logHDR* formats are stored in floating point values, whereas images in the other formats are stored as 8-bit integers. Examples of the formats can be seen in Fig. 7. Details on implementations and settings of the tone mappers we used in our experiments are listed in Table B.2 in Appendix B.

3.4. FP Detection and Evaluation Methodology

The six FP detectors described in Section 2.1 were applied on the sixteen image formats described in the previous section. This yielded 96 sets of FPs for each camera position. Since we wanted the FPs to be detected only in “meaningful” areas, a polygonal Region Of Interest (ROI) was defined in each image (see green polygons in Fig. 4 and Fig. 5). In the case of the 2D scene, the ROI covered the three posters but not the dot pattern. In the 3D scene, the ROI covered all the items placed on the desk, except the dot pattern again. The detectors were then applied in each ROI.

All ROIs have been divided into three non-overlapping areas (shadows, midtones and highlights, see Table 1), and the following evaluation was favoring image formats performing well in all three areas. This was motivated by the aim to evaluate the suitability of HDR imaging techniques for FP detection. In this context, it is desirable to detect features in *all* areas of a scene, regardless the illumination level.



Figure 7: Example images of image formats used in the experiments (the HDR and logHDR formats are not shown because they cannot be reproduced on paper)

Table 1: Partitioning of test scenes according to illumination levels.

Scene	Shadows (s)	Midtones (m)	Highlights (h)
2D	bottom-left poster	top-left poster	top-right poster
3D	statue of rhino	puppet, boxes	paper ball

Since every FP detector has a measure describing FP response/strength, two possible ways to assess the FP detectors exist:

1. **Constant threshold:** Specify a response threshold and let the detector find all the points stronger than the threshold, see Sec. 3.4.1.
2. **Constant number of FPs:** Specify the number of the strongest FPs to be detected, no matter how strong they are, see Sec. 3.4.2.

After computing responses of the FP detectors, we applied non-maximum suppression of the detector responses to get rid of clusters of FPs. This is important, because in an extreme case, where FPs are detected at every single pixel, the repeatability rate would reach 100%, which would make no sense. To avoid such a situation, we removed “weak” FPs by non-maxima suppression using a 21×21 pixels window. The size of the window was chosen with respect to the 9-pixel distance threshold for FP matching: 1 px the FP position + 9 px distance threshold left/up + 9 px distance threshold right/down + 1 px “safety” border around = 21 px window.

Details on implementations and settings of the FP detectors we used in our experiments are listed in Table B.1 in Appendix B.

3.4.1. Distribution of FPs

We used *constant threshold* to assess the *distribution* of FPs, i.e. how well and how many FPs each image format is able to present. See Table B.1 in Appendix B for the threshold values of individual FP detectors.

3.4.2. Repeatability of FPs

We used *constant number of FPs* to assess the *repeatability* of features. 300 FPs were detected in each image, except for those containing less features, e.g. due to overexposure. The repeatability of FPs was evaluated using the *Repeatability Rate* (RR).

To find out which FPs were repeated, we calibrated the cameras and established geometrical relations between individual views (details are in the next section). The matches between FPs were then established geometrically: a FP was considered as a match if it lay closer to its expected position than a distance threshold of 9 pixels. Note that *no FP descriptors were used* for the matching process.

RR was evaluated separately for areas of shadows (s), midtones (m), and highlights (h), which are described in Table 1. The values of RR in image pairs r_s , r_m , and r_h were calculated according to Eq. (7). The overall RR in an image pair is then the lower bound of the three values:

$$r = \min(r_s, r_m, r_h). \quad (11)$$

Similarly, the values of RR in image sequences in shadows, midtones, and highlights R_s , R_m , and R_h were calculated following Eq. (8). The overall RR in an image sequence is, again, the lower bound of the three values:

$$R = \min(R_s, R_m, R_h). \quad (12)$$

3.5. Geometric Matching of FPs

We performed an automatic camera calibration in each image sequence using the Camera Calibration Toolbox for Matlab⁴ and pattern-based camera pose estimation⁵. These calibrations provided the internal and external camera parameters at each camera position. Based on these parameters, we computed the geometric relations between individual views as described in [42].

In the case of the 2D scene, the geometric relation between the views is a planar homography and can be described by a 3×3 homography matrix \mathbf{H} . This matrix allows image coordinates to be mapped between two images so that:

$$\mathbf{x}_2 = \mathbf{H}_{12}\mathbf{x}_1 \quad (13)$$

where \mathbf{x}_1 and \mathbf{x}_2 are homogeneous image coordinates of corresponding points in first and second view and \mathbf{H}_{12} is a homography matrix describing planar homography between these two views. The positions of the detected FPs were matched fully automatically. A FP was considered found in an image if a point was detected in the radius of 9 pixels (which is less than 1% of the image size) from the predicted position, computed according to Eq. (13).

In the case of the general 3D scene, the views are related by the more complex concept of epipolar geometry. Without a 3D model of the scene, a 1:1 mapping between pixels is impossible. We solved this by manually annotating each view of the 3D scene sequences with a triangle mesh, see Fig. 8. The mesh contains 572 triangles which cover the same features in each view of the scene. Pixel coordinates inside the triangles were interpolated using barycentric coordinates. A FP was considered found in an image if a point was detected in the radius of 9 pixels from the predicted position, which was computed based on the triangle mesh.

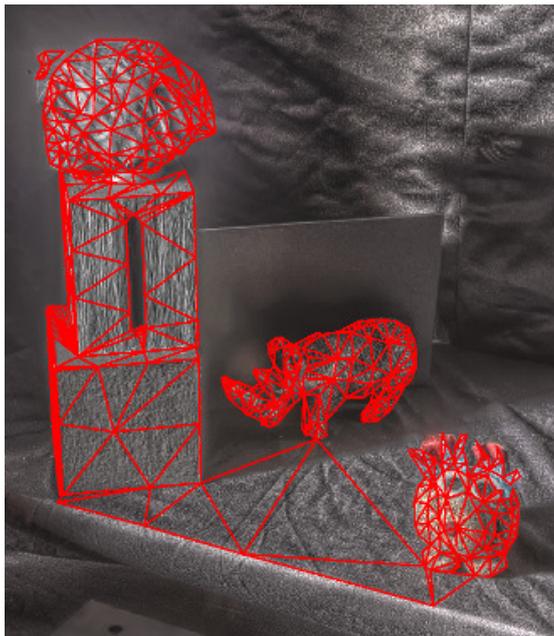


Figure 8: The triangle mesh with 572 triangles used to match FPs geometrically between individual views of the scene

4. Results

We evaluated the 2D and 3D scenes independently, because the previous work [4] indicates different behavior of FP detectors in these scenarios. In the design of experiments (described in

⁴Camera Calibration Toolbox for Matlab by Jean-Yves Bouguet: http://www.vision.caltech.edu/bouguetj/calib_doc/

⁵Automatic Camera Pose Estimation from Dot Pattern by George Vogiatzis and Carlo Hernández: <http://george-vogiatzis.org/calib/>

Section 3) we made an effort to stay as general as possible. Nevertheless, the results presented below inherently depend on the test images and scenes, and on settings of FP detectors and TMOs. Hence all the test images are made publicly available for download [43] and the settings of FP detectors and TM operators are documented in Appendix B.

The Shi-Tomasi FP detector had nearly identical results to the Harris corner detector, because the “cornerness” metric of Harris is an approximation of the metric of Shi-Tomasi, see Section 2.1. For the sake of clarity, we thus left the results of Shi-Tomasi out from further evaluation.

Applicability of any FP detector depends on (i) how many meaningful features does it detect and (ii) how stable those features are. Following this paradigm, we first assessed the distribution of FPs in individual image formats in Section 4.1. The stability of FPs was subsequently evaluated in Section 4.2.

4.1. Distribution of Feature Points

The first evaluated aspect was the distribution of FPs in the image. We set a *constant FP detector threshold* for all image formats for this purpose (i.e. we applied the first method described in Section 3.4). The distribution of FPs was assessed by visual inspection, which is the usual but inherently subjective method. We inspected three areas in each scene corresponding to shadows, midtones, and highlights, as defined in Table 1. The results are summarized in Tables 2 and 3.

Table 2: **The distribution of FPs** in individual image formats in the **2D scene**. “s” = shadows, “m” = midtones, and “h” = highlights. “0” = no detectable features in image, “-” = no FPs detected. “*”, “**”, and “***” = a few FPs, many FPs, and very many FPs detected, respectively. “÷” = FPs detected on noisy pixels.

2D	Harris			DoG			FastHes.			FAST			BRISK		
	s	m	h	s	m	h	s	m	h	s	m	h	s	m	h
HDR	-	**	**	-	***	***	-	***	***	-	**	**	-	**	**
logHDR	-	***	***	-	*	*	-	**	**	-	*	-	-	**	*
LDR	-	**	**	-	**	**	-	**	**	-	**	**	-	**	**
LDR.histeq	÷	*	**	-	*	**	-	*	**	-	*	**	-	*	**
LDR.CLAHE	-	**	**	-	***	**	-	***	**	-	**	**	-	**	**
gLogMap	-	**	**	-	-	-	-	-	-	-	-	-	-	-	-
gLarson97	-	*	**	-	*	**	-	*	**	-	*	**	-	*	**
gReinhard02	-	**	**	-	*	*	-	*	*	-	*	*	-	*	*
gReinhard05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
gMantiuk08	-	**	**	-	**	**	-	**	**	-	**	**	-	**	**
lReinhard02	-	**	**	-	**	**	-	**	**	-	**	**	-	**	**
lFattal02	-	**	**	**	**	***	**	***	***	*	**	**	**	**	**
lMantiuk06map	-	**	**	-	**	**	-	**	**	-	*	**	-	**	**
lMantiuk06equ	*	÷*	÷*÷	*	**	**	*	**	**	*	**	÷*	*	**	÷*
lFattal09	-	**	÷*÷	-	**	**	-	**	**	-	*	**	-	**	**

HDR. The assessment of FP distribution confirmed our assumption that FP detectors process the linear *HDR* images inefficiently. As expected, the FPs were detected in bright areas of the scenes only. More specifically, FP were detected in the highlights and the midtones (in the case of the 2D scene due to the lower dynamic range), and exclusively in the highlights (in the case of the 3D scene which has higher dynamic range). No FP were detected in the shadows, which is depicted in Tables 2 and 3 as the minus sign “-”.

Table 3: **The distribution of FPs** in individual image formats in the **3D scene**. “s” = shadows, “m” = midtones, and “h” = highlights. “0” = no detectable features in image, “-” = no FPs detected. “*”, “**”, and “***” = a few FPs, many FPs, and very many FPs detected, respectively. “÷” = FPs detected on noisy pixels.

3D	Harris			DoG			FastHes.			FAST			BRISK		
	s	m	h	s	m	h	s	m	h	s	m	h	s	m	h
HDR	-	-	**	-	-	**	-	-	**	-	-	*	-	-	**
logHDR	-	*	*	-	*	*	-	*	*	-	-	-	-	*	*
LDR	0	*	*	0	*	**	0	*	**	0	*	*	0	*	**
LDR.histeq	÷	÷*	÷	÷	÷*	÷	÷	÷*	÷	÷÷	÷*	÷	÷	÷*	÷
LDR.CLAHE	0	**	*	0	*	**	0	**	**	0	**	*	0	**	**
gLogMap	-	*	*	-	*	*	-	-	-	-	-	-	-	*	*
gLarson97	-	*	-	-	*	*	-	*	-	-	-	-	-	*	-
gReinhard02	-	*	-	-	*	*	-	*	-	-	-	-	-	*	*
gReinhard05	*	*	-	-	*	-	-	-	-	-	-	-	-	*	-
gMantiuk08	-	*	-	-	*	*	-	*	*	-	-	-	-	*	*
lReinhard02	-	*	*	*	*	*	-	*	-	-	*	*	-	*	*
lFattal02	**	**	*	***	***	***	***	***	***	**	***	*	***	***	**
lMantiuk06map	-	**	**	-	*	**	-	*	**	-	*	*	-	*	**
lMantiuk06equ	**	***	**	**	**	**	**	**	**	*	***	*	**	***	**
lFattal09	-	**	**	-	*	**	-	*	**	-	*	*	-	*	**

LogHDR. The logarithmically encoded HDR image yielded better FP distribution than the linear HDR. More FPs were usually detected in the midtones at the cost of less FPs detected in the highlights. This is more obvious in the case of 3D scene with its higher dynamic range, where the logarithmic mapping efficiently uses the logarithm function with higher base. In such a case, contrast in highlights gets more compressed and features are thus more attenuated. The base of the logarithm is proportional to the maximum luminance value in the image, see Eq. (10). Nevertheless, the total number of detected FPs was usually lower than in *HDR*, and still, no features were detected in the shadows. The FAST detector appeared to be sensitive to such compression, because it detected no FPs in the *logHDR* images of the 3D scene. It is interesting to note that Harris corner detector showed an opposite behavior than the other detectors: in the 2D scene, it detected more FPs in the *logHDR* format when compared to the *HDR*.

LDR. The *LDR* images (single exposures) naturally exhibited features only in well-exposed areas (i.e. midtones). The dark parts of the 2D scene were underexposed in the *LDR* images; no FPs were thus detected in the shadows, which can be observed in Table 2. The dark parts of the 3D scene were underexposed even more, leaving corresponding parts of the *LDR* images black. This is depicted in Table 3 as zeros “0”.

LDR + contrast enhancement. Contrast enhancing pre-processing techniques can help to detect more features in LDR images in a limited number of cases, depending on the presence of image noise. If large underexposed or overexposed noisy areas are present in the image, *histogram equalization* overly amplifies the noise. This has a destructive effect; it increases the number of detectable features, but they are in fact meaningless. This happened in the 2D scene when *Harris corner detector* detected five “feature points” on amplified noise in the *LDR.histeq* format, see the symbol “÷” in Table 2, and even more in the 3D scene when *all* FP detectors responded mainly to amplified noise, see Table 3. *CLAHE* did not amplify the noise so strongly (because the contrast boosting is limited), but its contribution to feature accentuation was also limited.

The *tone mapped* images, contrary to LDR, may contain features in all areas. However, significant differences between global and local TMOs were discovered.

Global TMOs. The global tone reproduction curves compress high dynamic range within the input image at the expense of the local contrast loss. Features thus get attenuated and are hard to detect. This behavior is more profound in the case of the 3D scene, which has a greater dynamic range than the 2D scene (see many “-” and only a few “★” in Table 3). The least detectable features are found in the *gReinhard05* image format. The FAST detector was, again, sensitive to the compression of contrast, because it detected almost no FPs in any globally tone mapped image of the 3D scene.

Local TMOs. The local tone mapping methods, on the other hand, generally excel in enhancement of local details. This effectively accentuates features, as can be seen in the bottom sections of Tables 2 and 3. However, features in the shadows are still harder to detect. The only tone mapping methods yielding detectable features in the shadows are *lFattal02* and *lMantiuk06equ*, although part of the features in *lMantiuk06equ* is merely amplified noise. Nevertheless, for the above reasons, *lFattal02* and *lMantiuk06equ* are the most suitable candidates for processing of HDR images for FP detection. This is an expected result, as *lFattal02* and *lMantiuk06equ* operate in gradient or contrast domain respectively to enhance image contrasts, and both are known to produce extremely detailed “hyper-realistic” HDR scene renditions. Nevertheless, the suitability has to be verified by the evaluation of the repeatability rate.

4.2. Repeatability Rate

The criterion of FP stability is the Repeatability Rate (RR). We let the tested FP detectors find the 300 strongest FPs in all examined image formats for this purpose (i.e. we applied the second method described in Section 3.4). RR was then evaluated separately for areas of shadows, midtones, and highlights, and the overall RR was obtained as a minimum of the three values, as defined in Eq. (11) and (12).

Previous studies evaluating FP detectors state that 3D scenes may yield lower RR than 2D scenes due to 3D phenomena such as occlusions, perspective, and clutter [4, 8]. This is, however, not the case in our study, because the 3D scene has no clutter and occlusions are very limited. Moreover, our study differs from the previous ones by the use of HDR scenes and imagery. The high dynamic range of light and its handling in images are more influential phenomena than the above mentioned 3D effects. This is even more emphasized by our definition of RR, which is evaluated separately in shadows, midtones, and highlights, and the minimum is taken as the overall repeatability. As a result, the average RR of the 3D scene was 16% while the RR of the 2D scene was only 9%. However, the average RR for individual combinations of scenes and image sequences differ significantly, as shown in Table 4.

Table 4: **Repeatability rate** of FPs in image sequences for individual scenes and image sequences according to Eq. (12), averaged over all FP detectors and image formats used.

	Distance	Lighting	Viewpoint	Average
2D	5 %	12 %	10 %	9 %
3D	23 %	5 %	20 %	16 %
Average	14 %	9 %	15 %	13 %

Scenes and image sequences. The most difficult setting is the lighting changing sequence of the 3D scene reaching an average RR of only 5% and having very poor RR both in shadows, mid-tones, and highlights. The difficulty is caused by changing shadows cast by the 3D structure, strongly altering the appearance of the scene. The distance changing sequence of the 2D scene also reaches an average RR of only 5%. However, in this case, it is only caused by poor repeatability of features in the shadows; RR of features in the midtones and highlights is comparable with other image sequences.

FP detectors. The average repeatability rate of individual FP detectors is shown in Table 5. Their performance reflects mainly their insensitiveness to image noise. While DoG and Fast Hessian are practically insensitive (an average RR of 25%), Harris and FAST are not (an average

Table 5: **Repeatability rate** of FPs in image sequences for individual FP detectors according to Eq. (12), averaged over both scenes, all image sequences, and all image formats used.

Harris	DoG	FastHes.	FAST	BRISK
20 %	25 %	25 %	19 %	5 %

RR of 20 % and 19 %, respectively). BRISK appeared to be very prone to image noise with the lowest average RR of only 5 %. Detailed graphs of RR in image pairs for each individual FP detector can be found in Appendix A. The RR for individual image formats is shown in Table 6.

HDR. The linear *HDR* yields an average RR of 0 %, which is caused by its poor performance in the shadows. Almost no useful features are detected, so they can hardly be repeated. This renders the linear *HDR* image format useless for contemporary FP detectors. However, we believe that future “HDR-aware” FP detectors may exploit its full potential, e.g. using adaptive thresholds.

LogHDR. The logarithmically transformed image (*logHDR*) performs much better (18 % on average), primarily thanks to its improved repeatability of FPs in the shadows. However, this is still not the best result among the tested image formats.

LDR, and contrast enhancement. The *LDR* format suffers from the absence of useful features in extreme shadows, and in highlights. This naturally causes an average RR of 0 %. The contrast enhancing pre-processing techniques (*LDR.histeq* and *LDR.CLAHE*) partially compensate this issue in a limited number of cases, as discussed in the previous section, but they also amplify potential image noise. This results in FPs at random positions, RR of which cannot be high: 9 % and 2 %, respectively.

Tone mapping. The tone mapped images yield better repeatability than HDR and LDR formats. The performance of all global TMOs is comparable (an average RR of 13 % – 24 %), while the performance of local TMOs is more diverse (an average RR of 1 % – 26 %).

The most similar results among the global TMOs are those of *gKiser12(t)* and *gReinhard02*, where the former is the *temporal* modification of the latter. See Table 6 and Fig. 9, where the results of the two image formats almost overlap. This indicates that the temporal coherency as implemented in *gKiser12(t)* has no significant effect on the stability of features in our setup.

Similarity can also be observed between the results of *gLarson97*, *gReinhard02*, and *gMantiuk08*. The average RR of features in these formats is, however, rather low (13 % – 15 %). This does not outperform even the 18% RR of *logHDR*.

The third and the second best average RR overall is achieved by *gLogMap* and *gReinhard05*, respectively. Please note that *gLogMap* is similar to *logHDR*, but it is succeeded by gamma correction. *gReinhard05*, on the other hand, is a tone mapper inspired by the human visual system, involving sigmoidal nonlinearity. As shown in Tables 2 and 3 nearly no FPs were detected in images produced by *gReinhard05*. Nonetheless, its RR is high. This means that features in *gReinhard05* are hard to detect, but once detected, they are stable. This complies with the fact that *gReinhard05* yielded the highest RR in the shadows amongst all image formats (RR in the shadows R_s is usually the lowest one from the triplet R_s, R_m, R_h).

The image format yielding the best average RR of 26 % is the local tone mapper operating in gradient domain *lFattal02*. It is one of the two image formats with the best distribution of FPs; the other is *lMantiuk06equ*. However, the two formats yield different RR: 26 % and 20 %, respectively. This implies that *lFattal02* accentuates features in a stable manner while *lMantiuk06equ* does not, see Fig. 10. The edge-preserving multiscale TMO *lFattal09* suffers from the same problem, yielding low average RR as well (1 %).

According to the results of our evaluation, the most suitable image format for FP detection is *lFattal02*. It accentuates the features so that many FPs can be detected in the images and simultaneously the features are stable. Generally, a TMO tailored for current FP detectors should obey two requirements: (i) Apply a non-linear (e.g., log or sigmoid) transformation to the luminance followed by gamma correction to make the image display-referred, and (ii) enhance local contrast while maintaining the noise level low to accentuate features.

Table 6: **Repeatability rate of FPs in image sequences** according to Eq. (12), averaged over all FP detectors used. Small gray numbers depict repeatability rate for shadows, midtones, and highlights (the % sign is left out for better readability). The resulting repeatability rate is the minimum of each triplet.

	2D			3D			Avg.
	Dist.	Light.	Viewp.	Dist.	Light.	Viewp.	
HDR	0 % 0 38 37	0 % 0 45 49	0 % 0 50 46	1 % 1 21 57	0 % 0 8 25	0 % 0 26 51	0 %
logHDR	6 % 6 33 37	18 % 18 59 62	8 % 8 46 42	37 % 37 37 45	4 % 4 14 15	33 % 34 33 47	18 %
LDR	0 % 0 33 35	1 % 1 47 50	0 % 0 47 44	0 % 0 36 43	1 % 1 16 10	0 % 0 41 60	0 %
LDR.histeq	4 % 4 36 38	3 % 3 39 43	2 % 2 54 53	30 % 30 34 34	6 % 8 6 8	9 % 9 30 17	9 %
LDR.CLAHE	1 % 1 39 22	2 % 2 51 39	8 % 8 49 35	2 % 2 30 37	2 % 2 16 11	0 % 0 38 56	2 %
gLogMap	12 % 12 33 35	28 % 28 61 51	18 % 18 43 41	32 % 46 32 35	11 % 11 13 14	30 % 38 30 41	22 %
gLarson97	1 % 1 36 37	9 % 9 48 55	4 % 4 50 44	35 % 38 38 35	8 % 8 11 9	22 % 39 27 22	13 %
gReinhard02	4 % 4 30 36	9 % 9 56 60	4 % 4 45 41	39 % 40 39 39	8 % 8 14 12	28 % 38 36 28	15 %
gKiser12(t)	3 % 3 29 36	11 % 11 59 53	3 % 3 45 41	37 % 40 39 37	7 % 7 14 11	28 % 38 36 28	15 %
gReinhard05	21 % 31 25 21	30 % 35 30 37	31 % 47 37 31	21 % 46 34 21	10 % 12 15 10	31 % 41 31 32	24 %
gMantiuk08	2 % 2 35 40	10 % 10 53 62	5 % 5 47 46	35 % 35 38 38	4 % 4 14 13	35 % 37 35 37	15 %
lReinhard02	1 % 1 32 37	7 % 7 52 59	6 % 6 46 43	25 % 25 37 37	3 % 3 13 13	33 % 42 33 35	13 %
lFattal02	16 % 16 36 36	34 % 34 61 62	38 % 38 48 45	29 % 47 29 39	12 % 12 13 15	28 % 38 28 45	26 %
lMantiuk06map	2 % 2 34 37	8 % 8 49 60	7 % 7 43 44	19 % 19 39 49	2 % 2 15 15	18 % 18 35 44	9 %
lMantiuk06equ	12 % 12 31 29	24 % 24 39 49	21 % 27 33 21	30 % 32 30 40	6 % 6 13 13	27 % 27 28 36	20 %
lFattal09	0 % 0 25 23	1 % 1 31 39	0 % 0 27 26	5 % 5 20 37	0 % 0 9 12	1 % 1 12 26	1 %

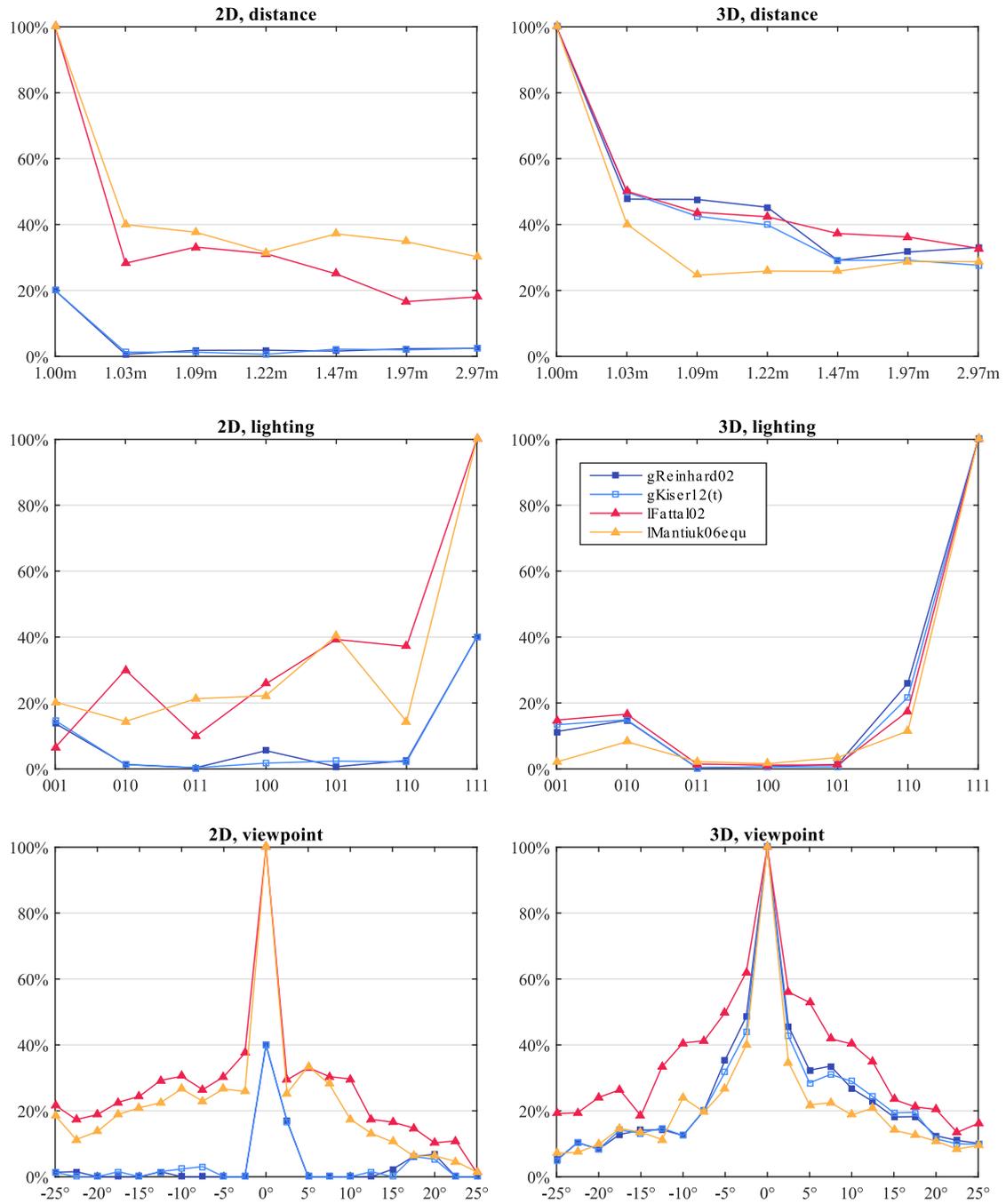


Figure 9: **Repeatability rate of FPs with respect to the reference image** according to Eq. (11), averaged over all FP detectors used. **(left)** 2D scene, and **(right)** 3D scene. **(top)** Distance changing sequence, **(middle)** lighting changing sequence, and **(bottom)** viewpoint changing sequence. RR in four different image formats is plotted with different colors

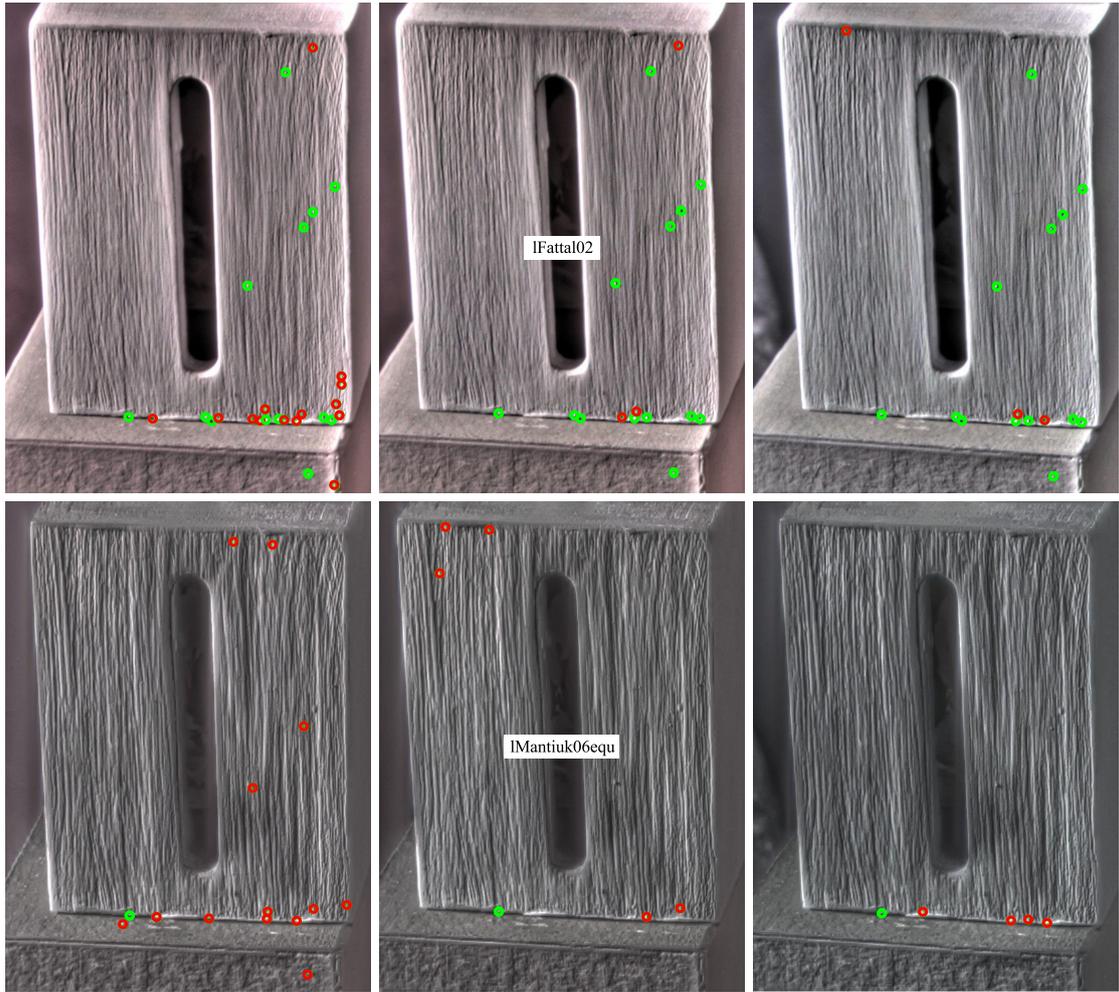


Figure 10: **Difference in repeatability** of FPs for the image formats IFattal02 (**top**) and IMantiuk06equ (**bottom**). Parts of consecutive images from the viewpoint changing sequence of the 3D scene (from left to right: 17.5° , 20.0° , 22.5°). Visible is the front part of the upper box. Stable FPs are marked as green circles (detected in all three images), unstable FPs are marked as red diamonds

5. Conclusions

This paper presents an evaluation of five different FP detectors using sixteen different image formats capturing scenes with a high dynamic range of luminance. High dynamic range in the test environments was achieved by illuminating a planar and a 3D scene with powerful tungsten lights in a totally dark room. The scenes were captured in three image sequences with varying camera viewpoints, camera distances, and scene lighting. The images of the scenes were processed in various image formats including HDR images, single exposure LDR images and tone mapped HDR images. The aim of the study was to evaluate the suitability of the image formats for FP detection. The evaluation was based on the investigation of the *distribution* and *repeatability rate* of FPs with respect to the image formats used.

Our results confirm the assumption that unprocessed linear HDR images are not suitable for current FP detectors because features are detected only in bright areas of images. This is due to the design of contemporary FP detectors, which assume a display-referred image as their input, and compute some kind of a derivative threshold. This strategy is inefficient in a scene-referred linear HDR image. This finding opens a promising avenue for future research into algorithms able to deal directly with HDR images, e.g. through the use of adaptive thresholds.

We investigated the possible work-around, using tone mapping to convert HDR images into standard display-referred LDR images. The global TMOs usually yield moderate repeatability, but the number of detected features is rather low due to the extensive compression of local contrast. This is disruptive for some FP detectors, e.g. for FAST. The local TMO, on the other hand, yield many features because the contrast is enhanced locally. However, big differences in repeatability between individual local TMOs exist, depending on how they accentuate features (and possibly also noise).

Our findings suggest to use a local TMO which, firstly, applies a non-linear transformation to the luminance (e.g., log or sigmoid) followed by gamma correction to make the image display-referred, and, secondly, enhances local contrast to accentuate features while maintaining the noise level low. This is important when using a FP detector sensitive to noise (e.g. Harris corner detector, FAST, and particularly BRISK). Our results indicate that the TMO method of choice may be *lFattal02* [33].

Future work will involve comparison of FP *descriptors*, which are often used together with FP detectors in practice.

Acknowledgments

This work was supported by the Technology Agency of the Czech Republic by project TE01020415 “V3C” and by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II) project “IT4Innovations excellence in science - LQ1602”. The research was further supported by SoMoPro II grant (financial contribution from the EU 7 FP People Programme Marie Curie Actions, REA 291782, and from the South Moravian Region). The work was also partially supported by the EU ICT COST Action IC1005 “HDRi: The digital capture, storage, transmission and display of real-world lighting”. Alan Chalmers is supported by a Royal Society Industrial Fellowship.

Our thanks also go to our WMG colleagues Elmedin Selmanovic, Jassim Happa, Thomas Bashford-Rogers, and Samuel Staton for their comments and support, to Tomáš Zemčík and Roman Tábi for annotating the images of 3D scene, and to Rossella Suma for measuring the values of light.

The images depicted on posters used in the 2D scene were downloaded from the HD Wallpapers website⁶ and may be used for non-commercial purposes only.

The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

⁶<http://www.hdwallpapers.in>

Appendix A. Extended Results

Detailed charts of repeatability rate of individual FP detectors are depicted in Fig. A.1 – A.5.

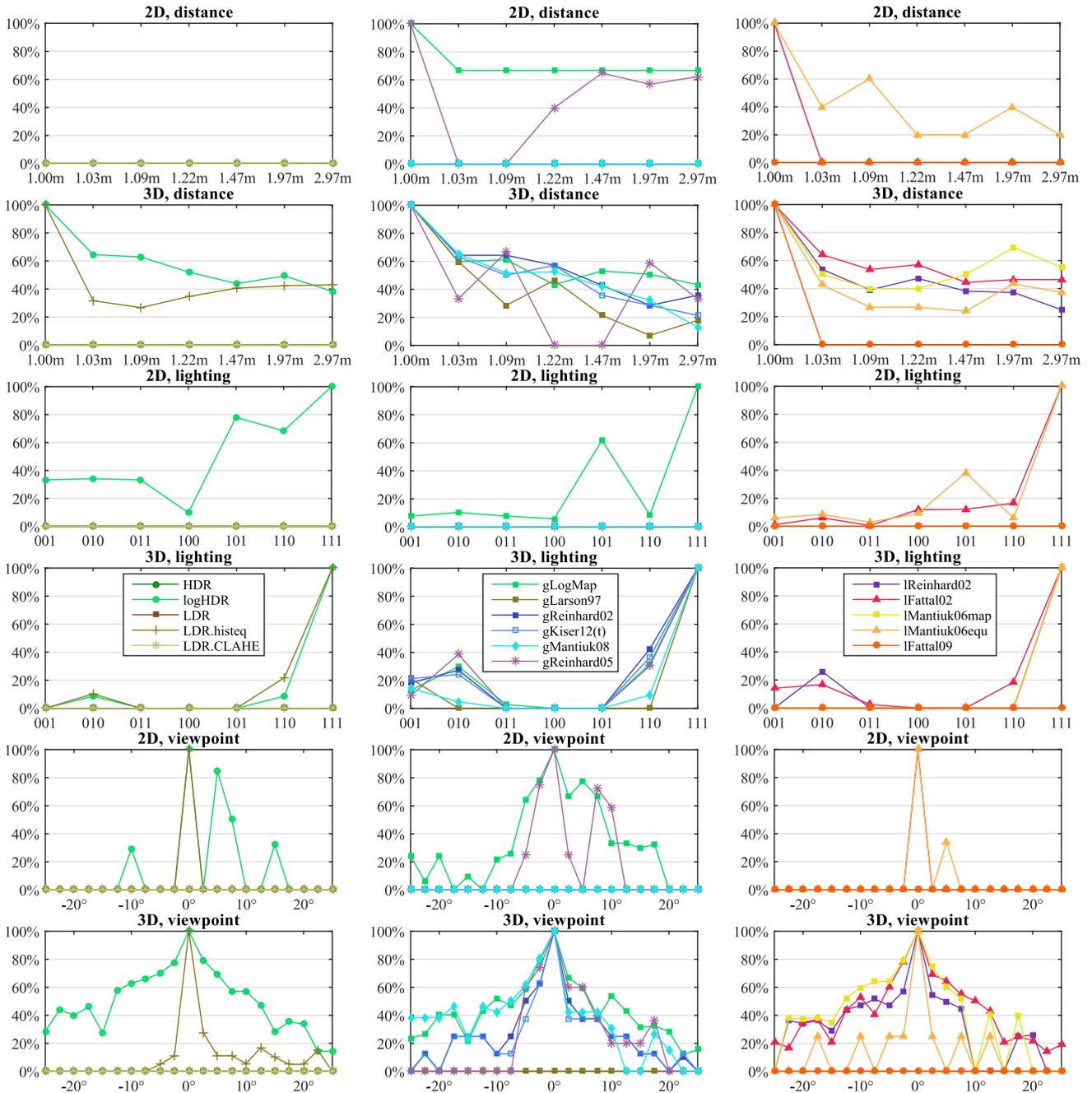


Figure A.1: Repeatability rate of the **Harris** FP detector with respect to the reference image, according to Eq. (11). (**left**) HDR and LDR, (**center**) global TMOs, (**right**) local TMOs. (**top**) distance changing sequence, (**middle**) lighting changing sequence, (**bottom**) viewpoint changing sequence for 2D and 3D scenes

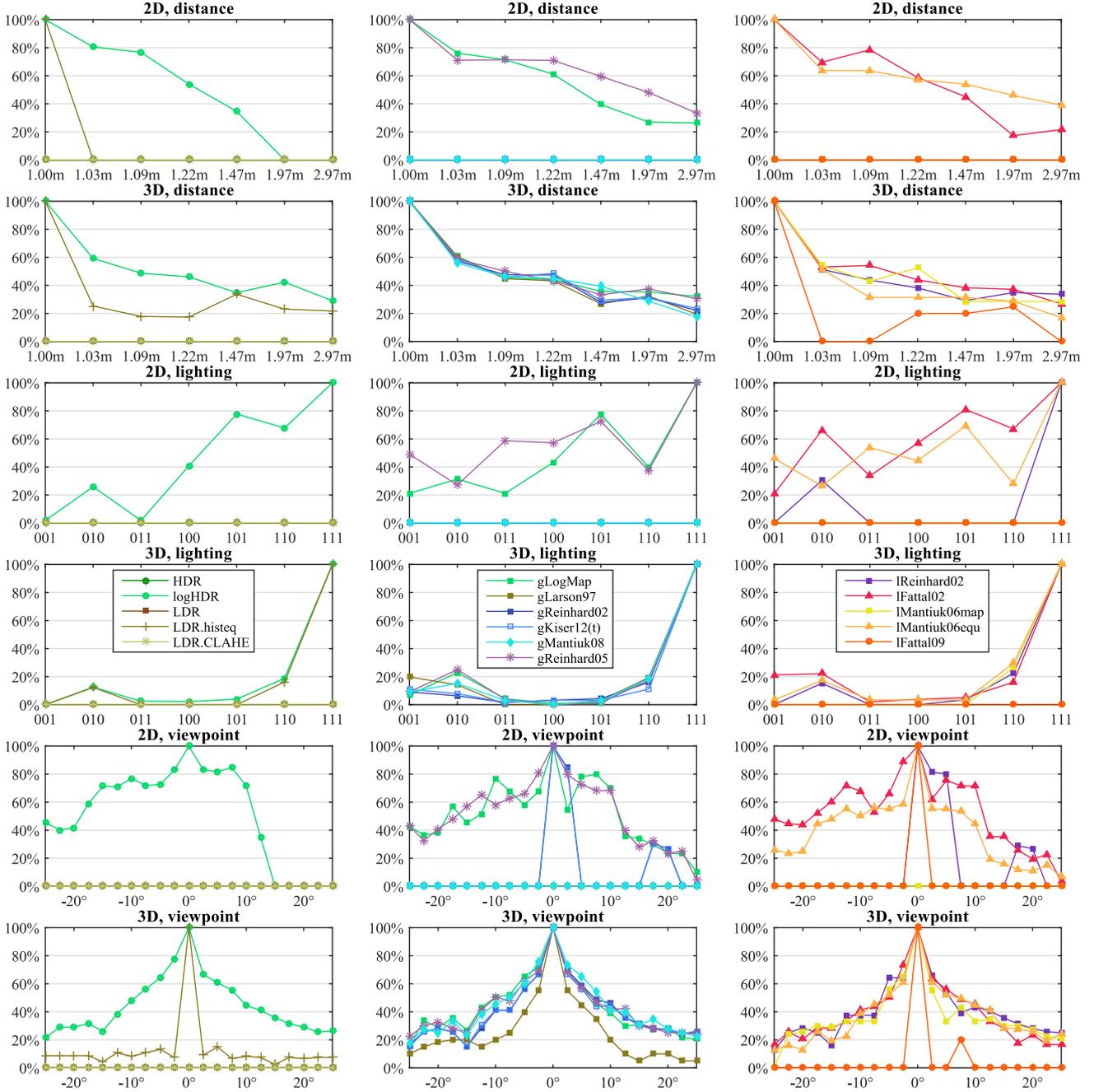


Figure A.2: Repeatability rate of the DoG FP detector with respect to the reference image, according to Eq. (11). (left) HDR and LDR, (center) global TMOs, (right) local TMOs. (top) distance changing sequence, (middle) lighting changing sequence, (bottom) viewpoint changing sequence for 2D and 3D scenes

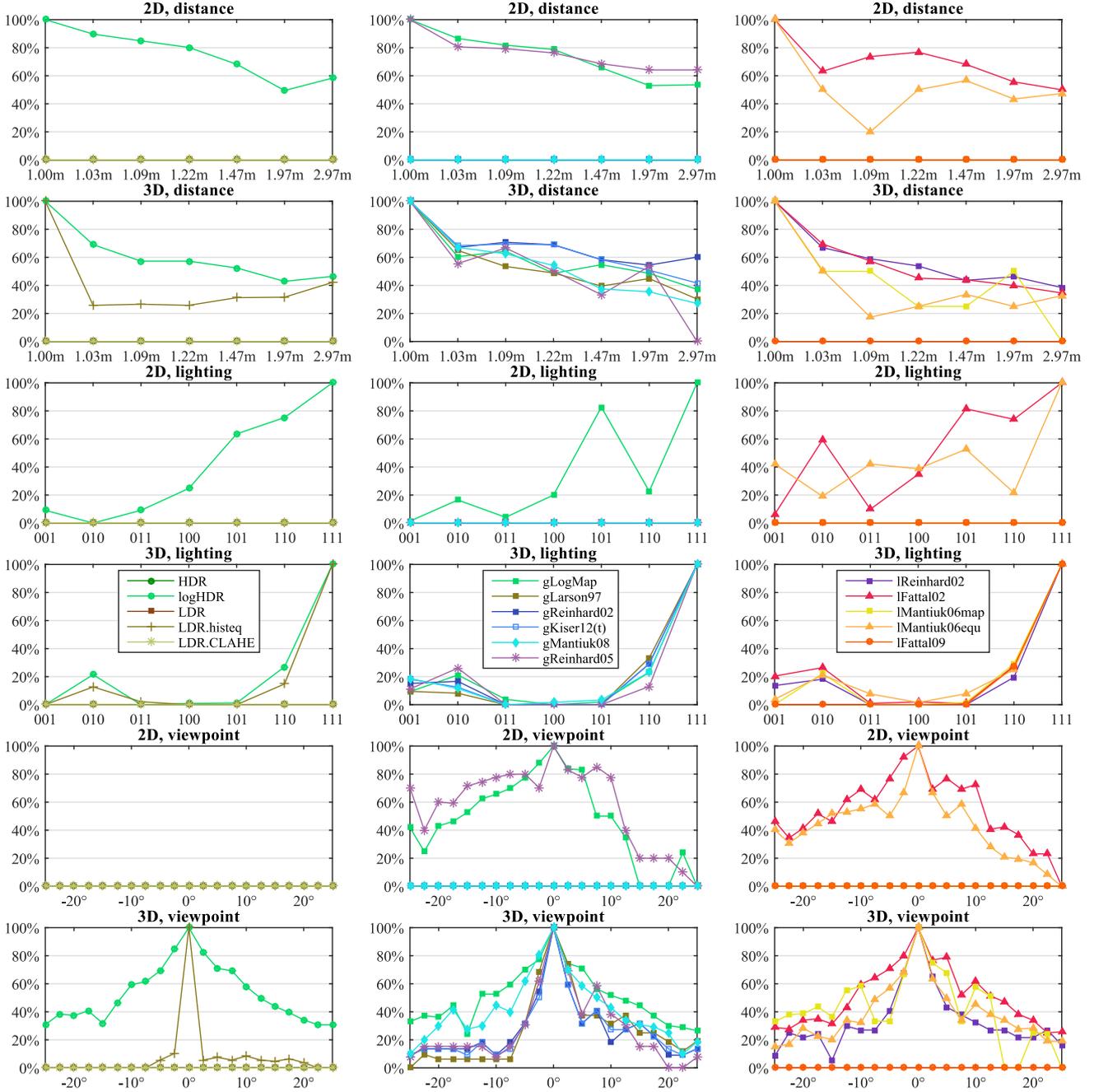


Figure A.3: Repeatability rate of the **Fast Hessian** FP detector with respect to the reference image, according to Eq. (11). (left) HDR and LDR, (center) global TMOs, (right) local TMOs. (top) distance changing sequence, (middle) lighting changing sequence, (bottom) viewpoint changing sequence for 2D and 3D scenes

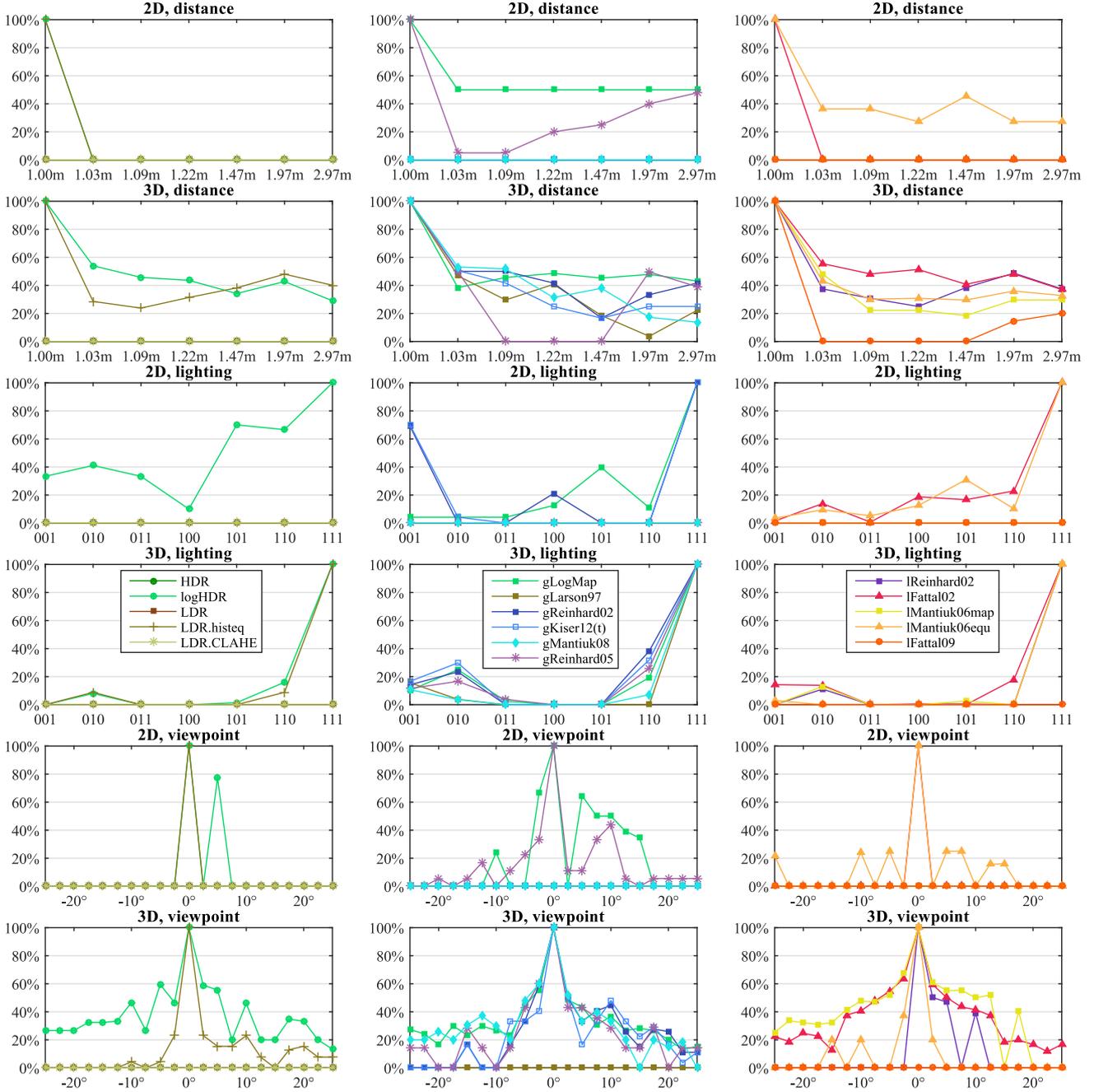


Figure A.4: Repeatability rate of the **FAST** FP detector with respect to the reference image, according to Eq. (11). (**left**) HDR and LDR, (**center**) global TMOs, (**right**) local TMOs. (**top**) distance changing sequence, (**middle**) lighting changing sequence, (**bottom**) viewpoint changing sequence for 2D and 3D scenes

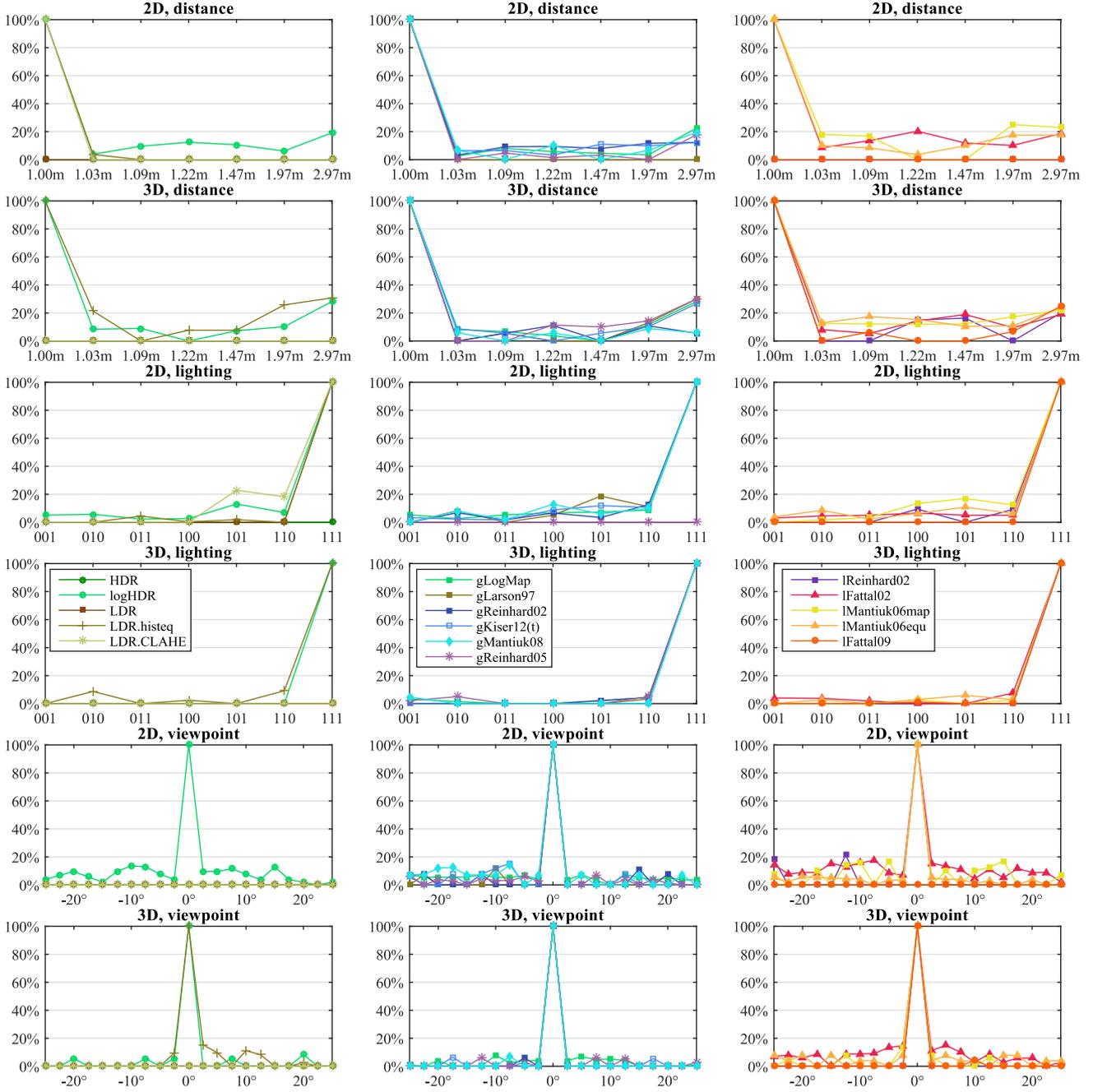


Figure A.5: Repeatability rate of the **BRISK** FP detector with respect to the reference image, according to Eq. (11). (**left**) HDR and LDR, (**center**) global TMOs, (**right**) local TMOs. (**top**) distance changing sequence, (**middle**) lighting changing sequence, (**bottom**) viewpoint changing sequence for 2D and 3D scenes

Appendix B. Settings of FP Detectors and TMOs

Table B.1 lists the parameter values of FP detectors, and Table B.2 lists the parameter values of tone mapping operators used in the experiments.

Table B.1: Parameter values of FP detectors used in our experiments. Parameters *not listed here* were set to *default values*.

FP Detector	Implementation	Parameter	Value
Harris	Matlab R2014b <code>detectHarrisFeatures()</code>	MinQuality	0.05
Shi-Tomasi	Matlab R2014b <code>detectMinEigenFeatures()</code>	MinQuality	0.2
DoG	VLFeat library 0.9.20 [44] <code>vl_sift()</code>	PeakThresh EdgeThresh	8 5
Fast Hessian	Matlab R2014b <code>detectSURFFeatures()</code>	MetricThreshold	3000
FAST	Matlab R2014b <code>detectFASTFeatures()</code>	MinContrast	0.2
BRISK	Matlab R2014b <code>detectBRISKFeatures()</code>	MinContrast	0.2

Table B.2: Parameter values of tone mapping operators used in our experiments. Parameters *not listed here* were set to *default values*.

TM Operator	Implementation	Parameter	Value	Gamma
gLogMap	HDR Toolbox for Matlab 1.0.7 LogarithmicTMO()			2.2
gLarson97	HDR Toolbox for Matlab 1.0.7 WardHistAdjTMO()			2.2
gReinhard02	HDR Toolbox for Matlab 1.0.7 ReinhardTMO()			2.2
gReinhard05	PFS tools 2.0.3 pfstmo_reinhard05			1.6
gMantiuk08	PFS tools 2.0.3 pfstmo_mantiuk08			-
lReinhard02	HDR Toolbox for Matlab 1.0.7 ReinhardTMO()	pAlpha pWhite pLocal pPhi	-1 -1 true 6.0	2.2
lFattal02	PFS tools 2.0.3 pfstmo_fattal02			-
lMantiuk06map	PFS tools 2.0.3 pfstmo_mantiuk06			2.2
lMantiuk06equ	PFS tools 2.0.3 pfstmo_mantiuk06	equalize-contrast	1.0	2.2
lFattal09	Fattal's code [45] EAW(), iEAW()	wavelet_type dist_func sigma	1 1 1	4.0

References

- [1] C. Schmid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors, *International Journal of computer vision* 37 (2) (2000) 151–172. doi:10.1023/A:1008199403446.
- [2] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Gool, A comparison of affine region detectors, *International Journal of Computer Vision* 65 (1) (2005) 43–72. doi:10.1007/s11263-005-3848-x.
- [3] F. Banterle, A. Artusi, K. Debattista, A. Chalmers, *Advanced High Dynamic Range Imaging: Theory and Practice*, AK Peters (CRC Press), Natick, MA, USA, 2011. doi:10.1201/b11373-5.
- [4] F. Fraundorfer, H. Bischof, A novel performance evaluation method of local detectors on non-planar scenes, in: *IEEE Conference on Computer Vision and Pattern Recognition – Workshops*, IEEE, 2005, p. 33. doi:10.1109/CVPR.2005.393.
- [5] V. Rodehorst, A. Koschan, Comparison and evaluation of feature point detectors, in: *5th International Symposium Turkish-German Joint Geodetic Days*, 2006.
- [6] P. Moreels, P. Perona, Evaluation of features detectors and descriptors based on 3d objects, *International Journal of Computer Vision* 73 (3) (2007) 263–284. doi:10.1007/s11263-006-9967-1.
- [7] I. Jazayeri, C. Fraser, Interest operators in close-range object reconstruction, *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 47 (2008) 69–74.
- [8] A. Gil, O. Mozos, M. Ballesta, O. Reinoso, A comparative evaluation of interest point detectors and local descriptors for visual slam, *Machine Vision and Applications* 21 (6) (2010) 905–920. doi:10.1007/s00138-009-0195-x.
- [9] I. Jazayeri, C. Fraser, Interest operators for feature-based matching in close range photogrammetry, *The Photogrammetric Record* 25 (129) (2010) 24–41. doi:10.1111/j.1477-9730.2009.00559.x.
- [10] S. Gauglitz, T. Höllerer, M. Turk, Evaluation of interest point detectors and feature descriptors for visual tracking, *International Journal of Computer Vision* 94 (3) (2011) 335–360. doi:10.1007/s11263-011-0431-5.
- [11] Y. Cui, A. Pagani, D. Stricker, Robust point matching in hdri through estimation of illumination distribution, *Pattern Recognition* (2011) 226–235doi:10.1007/978-3-642-23123-0_23.
- [12] G. Kontogianni, E. Stathopoulou, A. Georgopoulos, A. Doulamis, Hdr imaging for feature detection on detailed architectural scenes, *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2015) 325–330doi:10.5194/isprsarchives-XL-5-W4-325-2015.
- [13] Y. Li, S. Wang, Q. Tian, X. Ding, A survey of recent advances in visual feature detection, *Neurocomputing* 149, Part B (2015) 736–751. doi:10.1016/j.neucom.2014.08.003.
- [14] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International journal of computer vision* 60 (2) (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
- [15] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, *Computer Vision–ECCV 2006* (2006) 404–417doi:10.1007/11744023_32.
- [16] S. Leutenegger, M. Chli, R. Y. Siegwart, Brisk: Binary robust invariant scalable keypoints, in: *IEEE International Conference on Computer Vision 2011*, IEEE, 2011, pp. 2548–2555. doi:10.1109/ICCV.2011.6126542.

- [17] C. Harris, M. Stephens, A combined corner and edge detector, in: Alvey Vision Conference, Vol. 15, Manchester, UK, 1988, p. 50. doi:10.5244/C.2.23.
- [18] J. Shi, C. Tomasi, Good features to track, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 1994, pp. 593–600. doi:10.1109/CVPR.1994.323794.
- [19] M. Brown, D. Lowe, Invariant features from interest point groups, in: British Machine Vision Conference, Vol. 21, 2002, pp. 656–665. doi:10.5244/C.16.23.
- [20] E. Rosten, T. Drummond, Fusing points and lines for high performance tracking, in: IEEE International Conference on Computer Vision, Vol. 2, IEEE, 2005, pp. 1508–1515. doi:10.1109/ICCV.2005.104.
- [21] S. Palmer, Vision Science: Photons to Phenomenology, A Bradford book, Bradford Book, 1999.
- [22] E. Rosten, R. Porter, T. Drummond, Faster and better: A machine learning approach to corner detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (1) (2010) 105–119. doi:10.1109/TPAMI.2008.275.
- [23] R. Mantiuk, K. Myszkowski, H.-P. Seidel, A perceptual framework for contrast processing of high dynamic range images, ACM Transactions on Applied Perception 3 (3) (2006) 286–308. doi:10.1145/1166087.1166095.
- [24] M. May, T. Morris, K. Markham, W. J. Crowther, M. J. Turner, Towards Object Recognition using HDR Video, Stereoscopic Depth Information and SIFT, in: W. Tang, J. Colloso (Eds.), Theory and Practice of Computer Graphics, The Eurographics Association, 2009. doi:10.2312/LocalChapterEvents/TPCG/TPCG09/165–168.
- [25] M. May, M. Turner, T. Morris, Faw for multi-exposure fusion features, in: Y.-S. Ho (Ed.), Advances in Image and Video Technology, Vol. 7087 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 289–300. doi:10.1007/978-3-642-25367-6_26.
- [26] M. Čadík, M. Wimmer, L. Neumann, A. Artusi, Evaluation of hdr tone mapping methods using essential perceptual attributes, Computers & Graphics 32 (3) (2008) 330–349. doi:10.1016/j.cag.2008.04.003.
- [27] J. Petit, R. Brémond, A. Tom, Evaluation of tone mapping operators in night-time virtual worlds, Virtual Reality (2012) 1–10doi:10.1007/s10055-012-0215-4.
- [28] G. Larson, H. Rushmeier, C. Piatko, A visibility matching tone reproduction operator for high dynamic range scenes, IEEE Transactions on Visualization and Computer Graphics 3 (4) (1997) 291–306. doi:10.1109/2945.646233.
- [29] W. B. Green, Digital image processing: a systems approach, Electrical-Computer Science and Engineering Series, Van Nostrand Reinhold Co., 1983.
- [30] E. Reinhard, M. Stark, P. Shirley, J. Ferwerda, Photographic tone reproduction for digital images, ACM Transactions on Graphics 21 (3) (2002) 267–276. doi:10.1145/566654.566575.
- [31] E. Reinhard, K. Devlin, Dynamic range reduction inspired by photoreceptor physiology, IEEE Transactions on Visualization and Computer Graphics 11 (1) (2005) 13–24. doi:10.1109/TVCG.2005.9.
- [32] R. Mantiuk, S. Daly, L. Kerofsky, Display adaptive tone mapping, in: ACM Transactions on Graphics, Vol. 27, ACM, 2008, pp. 68:1–68:10. doi:10.1145/1360612.1360667.
- [33] R. Fattal, D. Lischinski, M. Werman, Gradient domain high dynamic range compression, ACM Transactions on Graphics 21 (3) (2002) 249–256. doi:10.1145/566654.566573.
- [34] R. Fattal, Edge-avoiding wavelets and their applications, ACM Transactions on Graphics 28 (3) (2009) 22. doi:10.1145/1531326.1531328.

- [35] C. Kiser, E. Reinhard, M. Tocci, N. Tocci, Real time automated tone mapping system for hdr video, in: IEEE International Conference on Image Processing, 2012, pp. 2749–2752.
- [36] R. Yates, R. Lyons, Dc blocker algorithms [dsp tips&tricks], IEEE Signal Processing Magazine 25 (2) (2008) 132 – 134. doi:10.1109/MSP.2007.914713.
- [37] E. Reinhard, W. Heidrich, P. E. Debevec, S. Pattanaik, G. Ward, K. Myszkowski, High dynamic range imaging: acquisition, display, and image-based lighting, Academic Press, 2010.
- [38] P. E. Debevec, J. Malik, Recovering high dynamic range radiance maps from photographs, in: SIGGRAPH, 1997, p. 369. doi:10.1145/258734.258884.
- [39] K. Wallis, Seasonal adjustment and relations between variables, Journal of the American Statistical Association 69 (345) (1974) 18–31. doi:10.1080/01621459.1974.10480123.
- [40] B. Příbyl, A. Chalmers, P. Zemčík, Feature point detection under extreme lighting conditions, in: Spring Conference on Computer Graphics, Comenius University in Bratislava, 2012, pp. 156–163. doi:10.1145/2448531.2448550.
- [41] K. Zuiderveld, Contrast limited adaptive histogram equalization, in: Graphics gems IV, Academic Press Professional, Inc., San Diego, CA, USA, 1994, pp. 474–485.
- [42] R. Hartley, A. Zisserman, Multiple view geometry in computer vision, 2nd Edition, Cambridge University Press, 2004. doi:10.2277/0511188951.
- [43] B. Příbyl, A. Chalmers, P. Zemčík, L. Hooberman, M. Čadík, Supplementary data for the research on feature point detection in high dynamic range images, <http://www.fit.vutbr.cz/~ipribyl/FPinHDR/> (December 2015).
- [44] A. Vedaldi, B. Fulkerson, VLFeat: An open and portable library of computer vision algorithms, <http://www.vlfeat.org/> (2008).
- [45] R. Fattal, Edge-avoiding wavelets and their applications, <http://www.cs.huji.ac.il/~raananf/projects/eaw/> (September 2012).