

Feature Point Detection under Extreme Lighting Conditions

Bronislav Příbyl*
Faculty of Information Technology
Brno University of Technology

Alan Chalmers†
International Digital Laboratory, WMG
University of Warwick, UK

Pavel Zemčík‡
Faculty of Information Technology
Brno University of Technology

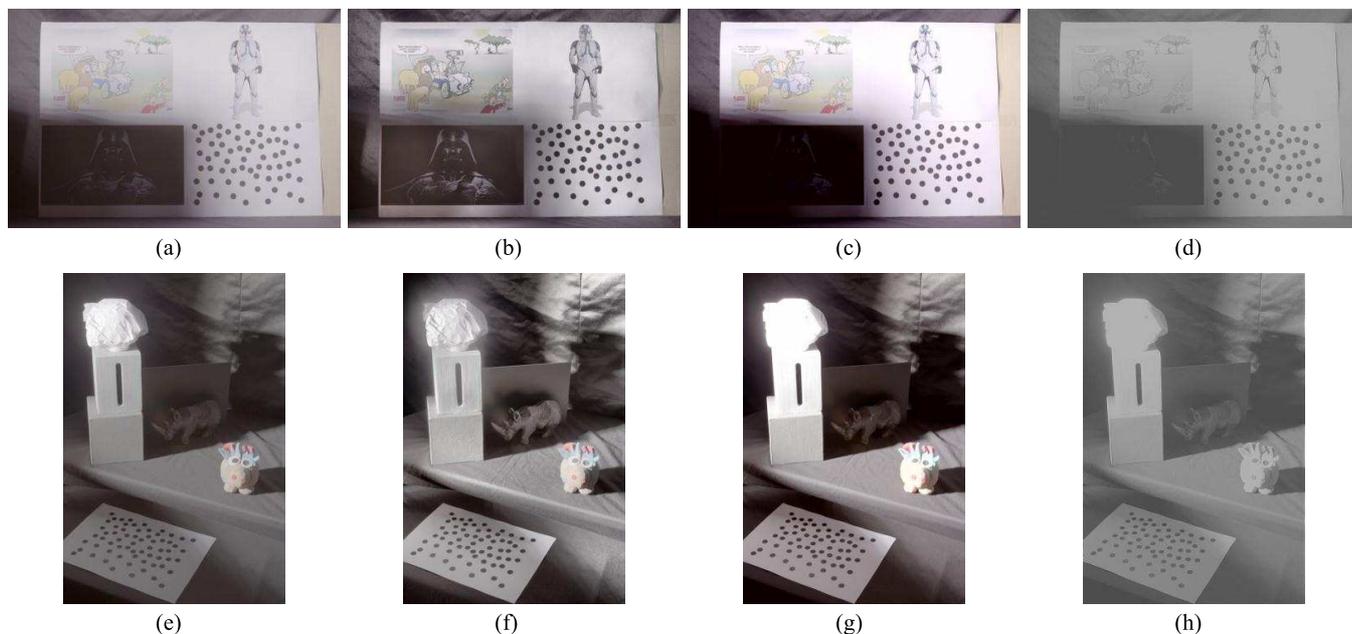


Figure 1: Example images from sequences used in the experiments. A 2D scene (containing just one plane; top row) and a 3D scene (containing objects of different shapes; bottom row) were captured using four different image formats. These are, from left to right: HDR image tone mapped using a global tone mapper: GTM format (a, e). HDR image tone mapped using a local tone mapper: LTM format (b, f). Ordinary low dynamic range image: LDR format (c, g). LDR image filtered using Wallis filter: WAL format (d, h).

Abstract

This paper evaluates the suitability of High Dynamic Range (HDR) imaging techniques for feature point detection under extreme lighting conditions. The conditions are extreme in respect to the dynamic range of the lighting within the test scenes used. This dynamic range cannot be captured using standard low dynamic range imagery techniques without loss of detail. Four widely used feature point detectors are used in the experiments: Harris corner detector, Shi-Tomasi, FAST and Fast Hessian. Their repeatability rate is studied under changes of camera viewpoint, camera distance and scene lighting with respect to the image formats used. The results of the experiments show that HDR imaging techniques improve the repeatability rate of feature point detectors significantly.

CR Categories: I.4.6 [Computing Methodologies]: Image Processing and Computer Vision—Segmentation

Keywords: Feature Point Detection, Interest Point Detection, Corner Point Detection, Harris Corner Detector, Shi-Tomasi, FAST, Fast Hessian, SURF, High Dynamic Range Imagery, HDR, Wallis Filter, Tone Mapping

*e-mail: ipribyl@fit.vutbr.cz

†e-mail: alan.chalmers@warwick.ac.uk

‡e-mail: zemcik@fit.vutbr.cz

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 the image [Schmid et al. 2000]. These features mainly consist of blobs, edges or points. In the case of points, these are referred to as corner points, interest points or most often the “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.

The detection of FPs is strongly dependent on the illumination of the scene at the moment of image capture [Mikolajczyk et al. 2005]. Demanding lighting conditions or wrong camera settings can cause FP detectors to fail to detect any points. When capturing images under such conditions, one has to carefully set the camera and arrange the scene (which is a limiting factor and sometimes cannot be done at all). An alternative

©ACM, 2012. This is the authors version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Proceedings of the 28th Spring Conference on Computer Graphics, May 2012. <http://doi.acm.org/10.1145/2448531.2448550>

approach is to use High Dynamic Range (HDR) imagery.

HDR imagery captures and stores information about the full amount of light in a scene rather than the perceived colour in an image, which is stored in traditional Low Dynamic Range (LDR) imagery. This difference between HDR and traditional LDR technology is the use, by HDR, of floating point numbers capable of representing a theoretically infinite light intensity range, rather than 8-bit integer values of LDR that limit the intensity range to 0-255. The HDR representation provides more detailed information about a scene and thus has the potential to improve the performance of computer vision tasks. This is particularly true if dealing with images of the natural world where the average luminance levels may vary approximately between 10^{-3} cd/m² for a starlit night and 10^6 cd/m² on a sunny day [Banterle et al. 2011]. This difference between the luminance levels demonstrates the necessity of using HDR imagery to capture the full range of light in the real world.

There have been many comparisons and evaluations of feature point detectors, both general and application specific ones. To the best of our knowledge, all these evaluations have been carried out using classical LDR images only. There are a few recent papers that consider HDR imagery in FP detection (e.g. [Cui et al. 2011]), but no comparison with LDR has been done. This paper investigates the impact of HDR imagery, compared to LDR, to significantly improve the performance of FP detectors.

The rest of this paper is organized as follows: In section 2, we briefly describe the selected four FP detectors for our experiments, and discuss the literature on previous comparisons of FP detectors. Possible image pre-processing steps are also discussed. Section 3 details the setup for subsequent evaluation. The results are presented and analysed in section 4. Finally, we conclude our work and make suggestions for future work in section 5.

2 Related Work

2.1 Feature Point Detectors

Although a number of feature point detectors have been proposed in prior literature, the following four have been shown to outperform other detectors [Jazayeri and Fraser 2008; Gil et al. 2010; Gauglitz et al. 2011]. They are thus widely used in many applications.

Harris Corner Detector: This method is based on the local auto-correlation function reflecting local intensity changes in the image [Harris and Stephens 1988]. 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 pixel intensity in the x and y directions at point \mathbf{x} . Then the point score $R(\mathbf{x})$ is computed as

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

where the determinant $\det(\mathbf{M})$ corresponds to the product of the two eigenvalues of \mathbf{M} and the trace $\text{tr}(\mathbf{M})$ to their sum, respectively. Associated eigenvectors represent two perpendicular directions of the greatest intensity change in the image. The constant k is a sensitivity factor.

Shi-Tomasi: The minimum eigenvalue detection method proposed by Shi and Tomasi [1994] relies on the same second moment matrix \mathbf{M} as the Harris detector does, but explicitly computes its eigenvalues contrary to Harris. This results in higher computational demands but also in feature points which are more stable for tracking.

FAST: The Features from Accelerated Segment Test method (or local intensity comparison) works on another principle [Rosten and Drummond 2005]. This method 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 which describes what features will be detected (both corners and edges or just corners). The circle considered has usually a radius of 3 pixels in practical applications. This yields 16 surrounding pixels to be tested. False candidate corner points can be eliminated quickly by testing the intensity criterion for surrounding pixels in a non-sequential manner, accelerating the whole procedure significantly.

Fast Hessian: This is the detector part of the so called SURF (Speeded up Robust Features [Bay et al. 2006]) combined feature detector and descriptor. In this paper, we only use the detection part. This detector approximates the Hessian matrix

$$\mathcal{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix} \quad (3)$$

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 $L_{yy}(\mathbf{x}, \sigma)$. The computational speed-up is achieved by usage of simple box filters instead of Gaussian second order partial derivatives filters and, subsequently, by use of precomputed integral images which allow the convolution of arbitrary size to be substituted by just 3 additions or subtractions.

2.2 Comparison of FP Detectors

There has been a number of previous papers on the comparison and evaluation of feature point detectors in the last decade. The first extensive study was published by Schmid et al. [2000]. The authors compared FP detectors on two planar scenes under changes in rotation, viewpoint and illumination, as well as with artificially added image noise. They were the first to introduce and evaluate the repeatability rate. The Harris corner detector was found to outperform the others, but since that time, many new FP detectors have been proposed. Mikolajczyk et al. [2005] compared affine-invariant detectors. They didn't focus just on point detectors but generally on region detectors. The scenes used for their experiments were also planar or near-planar. Fraundorfer and Bischof [2005] continued the work of Mikolajczyk et al. and extended it by introducing a new tracking method. By using this method, they were the first to be able to evaluate the detectors on non-planar scenes as well. Rodehorst and Koschan [2006] compared performance of detectors on a set of artificial planar scenes as well as natural 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 [2007] explored the performance of combinations of detectors and descriptors with a testbed of 3D objects. They exploited geometric constraints between triplets of views.

More recently, many application-specific studies of FP detectors have been published. Such studies naturally focus on certain parameters of the detectors. For example, Gil et al. [2010] evaluated behaviour of the detectors and descriptors on 2D and 3D scenes with respect to their use in visual navigation. They thus studied the stability of detected features throughout the whole image sequences, rather than just in couples of images. Jazayeri and Fraser [2010] concentrated on feature-based matching and close range photogrammetry. This paper thus focused on the detection rate of the detectors. They were also one of the first who evaluated the speed of the detectors. Gauglitz et al. [2011] compared the detectors and descriptors with regard to real-time visual tracking, thus concentrating on execution time and precision.

2.3 Image Pre-Processing

Some studies suggest a pre-processing step should be used to enhance local contrast when detecting features in an image, as low contrast can limit performance of the detectors. This is particularly important if variable imaging conditions occur. Jazayeri and Fraser [2010] and Zhang et al. [2009] proposed to use e.g. histogram equalization or Wallis filtering.

Wallis filter [Wallis 1974]:

Local contrast of an image is enhanced by adjusting pixel intensity in small regions to match predefined values of mean and standard deviation. The resulting image contains more detail even in areas with low and high brightness. It has been found that the detectors typically find more suitable FPs in images pre-processed with a Wallis filter [Ohdake and Chikatsu 2005; Remondino and Zhang 2006].

Local contrast enhancement is not needed when working directly with HDR imagery because it contains information about the true contrast of captured scene. However, when converting an HDR image to an 8-bit image (able to represent much lesser contrast) by a tone mapping operator, it is desirable to preserve the contrast as much as possible. Many tone mapping operators exist [Ledda et al. 2005] but it is beyond the scope of this work to evaluate them all. Therefore, just the basic ones are mentioned – one global and one local.

Global tone mapping: The operator consists of a \log_2 transformation of pixel intensities followed by a linear scaling and quantization into the integer interval $(0, 255)$.

Local tone mapping: The operator is based on contrast limited adaptive equalization of luminance histogram [Zuiderveld 1994]. The input image is divided into 8×8 tiles and the contrast of each tile is enhanced so its histogram is uniform. The neighbouring tiles are then combined using bi-linear interpolation to avoid artificial boundaries.

3 Experimental Setup

Two test scenes shown in Figure 1, have been created to evaluate the selected feature point detectors under extreme lighting conditions:

- A planar (2D) scene containing three different posters in A4 format next to each other, attached to a box.
- A 3D scene containing several non-planar rigid objects.

These scenes were placed into a totally dark room and illuminated by two 2kW reflectors to create the extreme lighting conditions. The 2D scene was made of a dark poster placed into a shadow, another poster lit by one reflector and a bright poster lit by both reflectors. The 3D scene was made analogously to the 2D scene: A coarse, dark statuette of a rhino placed into a shadow, a puppet lit by one reflector and a ball of creased paper lit by both reflectors. Both scenes have been designed to generate as much dynamic range of light as possible while containing textured areas feasible for detection of FPs. The 2D scene has an average dynamic range of 12.1 stops and the 3D scene, 13.7 stops. Both scenes also contain a sheet of paper with a dot pattern used for automatic calibration of internal and external parameters of the camera.

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 Figure 2.

Viewpoint changing sequence: The camera was moved following a circular trajectory with its centre in the scene with a step of 2.5° . Since the scenes were captured 21 times the total viewpoint range was 50° .

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 because the FP detectors are much more sensitive to scene distance changes at close range than at far range.

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 reflectors mentioned before. 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) reflector 1, (3) reflector 2. For example, the code “101” means that only the strip lights and the reflector 2 were on.

To capture all the dynamic range in the scenes, several 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. 16Mpx taken at one position. The multiple exposures were then combined into a single HDR image. The HDR image was subsequently tone mapped using the simple global and local tone mapper described in section 2.3. From the HDR image containing all the dynamic range, a “single exposure” LDR image has been made by clipping the dynamic range to 8 stops. Bounds of the clipping interval were chosen so that the number of over and underexposed pixels was minimised. This corresponds to an LDR image taken with ideal camera settings. As the Wallis filter is recognized as an image pre-processing operator feasible for FP detection, we also filtered the LDR images with the Wallis filter. The aim was to investigate if the Wallis filter improves FP detection under extreme lighting conditions more than tone mapping operators. All this resulted in scenes captured in four different image formats at each camera position. These are the following:

GTM: HDR image tone mapped with the global tone mapper.

LTM: HDR image tone mapped with the local tone mapper.

LDR: Low dynamic range image.

WAL: LDR image filtered with Wallis filter.



Figure 2: Example images from sequences used in the experiments. Viewpoint changing sequence: leftmost image (-25° , **a**), middle image (0° , **b**) and rightmost image ($+25^\circ$, **c**). Lighting changing sequence: lighting conditions “100” (**d**), “111” (**e**) and “010” (**f**). Distance changing sequence: nearest image (100 cm, **g**), further image (122 cm, **h**) and furthest image (297 cm, **i**). Magnified view of the puppet from the 3D scene (**j**). Feature points detected using the Fast Hessian detector are marked with green dots. The feature points were detected only inside of the Regions Of Interest (ROI) which are marked as green polygons in the reference images for each sequence (**b**, **e**, **g**). All displayed images are in LTM format.

The FP detectors have been tested on these four image formats, which examples can be seen in Figure 1. The detectors have not been tested directly on HDR images because their current implementations cannot cope with HDR imagery. This could be solved with new implementations capable of processing HDR images, but the risk of skewed results of the experiments would emerge. It was thus decided to run current implementations on 8-bit images only. As the four FP detectors were carried out on four image formats, they yielded 16 sets of feature points for each camera position. Since we wanted the FPs to be detected only in “meaningful” areas, a continuous Region Of Interest (ROI) was defined in each image (see green polygons in Figure 2). 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, excepting the dot pattern again. The detectors were then executed in each ROI.

Since every detector has a metric for feature points describing their response/strength, there are two ways in which to carry out the detectors: The first way is to specify a response threshold and let the detector find all the points stronger than the threshold. This way is suitable when working with single image format and allows the detection rate (i.e. how many feature points a given detector is able to generate in given image) to be evaluated. However, since there are four different image formats this approach was not suitable. Each image format yields a significantly different detector response for the same image regions as is illustrated in Figure 3, so the thresh-

olds would not match. Therefore a different approach is needed. It was thus chosen to specify the number of strongest feature points to be detected, which was set to 95-100 points in each image.

Finally, we performed automatic camera calibration in each image sequence using the Camera Calibration Toolbox for Matlab¹ and pattern-based camera pose estimation². These provided the internal and external camera parameters at each camera position. Based on these parameters, we computed the geometric relations between individual views [Hartley and Zisserman 2004]. In the case of a 2D scene, the relation 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

$$\mathbf{x}_2 = \mathbf{H}_{12}\mathbf{x}_1, \quad (4)$$

where \mathbf{x}_1 and \mathbf{x}_2 are homogeneous image coordinates of corresponding point in first and second view and \mathbf{H}_{12} is a homography matrix describing planar homography between these two views. In the case of a general 3D scene, the views are related by the more complicated concept of epipolar geometry. A direct map between two image points is no longer possible since an image point \mathbf{x}_1 in one view defines an epipolar line \mathbf{l}_{x_1} in the other view; Such a relation can be described by a 3×3 fundamental matrix \mathbf{F} :

$$\mathbf{l}_{x_1} = \mathbf{F}_{12}\mathbf{x}_1. \quad (5)$$

Since the dot pattern was not placed in the middle of the scenes, the geometric relations based on the automatic camera calibration further away from the pattern were not precise enough. The error of image-point positions was approximately tens of pixels. It was thus decided to manually mark 4 corresponding image points in each view of the 2D scene (e.g. the corners of the posters) and compute more precise homography matrices based on these markings. In addition, in the case of 3D scene, 7 corresponding image points were manually marked in each view (e.g. corners of the boxes or hair of the puppet) and more precise fundamental matrices were computed. These quantities allowed us to exploit geometric constraints between the views in the subsequent evaluation.

4 Evaluation

We evaluated the 2D and 3D scenes independently, as the behaviour of the FP detectors differs significantly between these cases (see Figure 4). The results presented in this paper are highly dependent on the sets of scenes and images used.

The first aspect evaluated was the distribution of FPs in the image. This was achieved by visual inspection which is the usual, but inherently subjective method. In general, all detectors tended to detect FPs on real (i.e. meaningful) features. The only detector slightly deviating from this was the FAST detector which also produced some clusters of random FPs on edges. This behaviour was caused by the default value 157.5° of the threshold angle (see section 2.1) and can be eliminated by setting a more acute angle value. In general, the distributions of FPs across the regions of interest were not uniform. There were some areas with lower contrast, mainly darker ones, where only a few FPs were detected. This issue was partially eliminated in image formats produced by local contrast enhancing

operators, i.e. the local tone mapper and Wallis filter. Over and underexposed areas will exist in LDR images when using them to capture scenes under extreme lighting conditions. These areas do not contain any detail and therefore no FPs can be detected. This problem does not arise when using HDR imagery.

The other and main criterion we evaluated was the Repeatability Rate (RR). This is currently considered to be standard metric of FP detectors. RR describes how FP detection is independent of imaging conditions [Schmid et al. 2000] and is defined as the ratio between the number of FPs detected in the image and the number of FPs detected in the reference image. It is desirable to achieve a repeatability rate as high as possible.

When deciding whether a feature point was detected or not in an image, the geometric constraints between the image and the reference image only were exploited. No FP descriptors were used. In the case of the 2D scene, the positions of the detected FPs were matched automatically. A FP was considered found in an image if there was a point detected in the radius of 35 px (which is less than 1 % of the image size) from predicted position, computed according to Equation 4. In the case of the 3D scene, candidate FPs were chosen as those lying closer than 35 px to the epipolar line computed according to Equation 5. Since there might be several candidate points next to each epipolar line, we also checked them manually to avoid false correspondences. The repeatability rate of tested detectors is shown in Figure 4 and Table 1.

Sequence	Scene	GTM	LTM	LDR	WAL
Viewpoint	2D	70 %	90 %	71 %	71 %
	3D	57 %	51 %	57 %	57 %
Lighting	2D	67 %	61 %	64 %	68 %
	3D	55 %	39 %	40 %	39 %
Distance	2D	85 %	92 %	85 %	85 %
	3D	59 %	50 %	52 %	49 %

Table 1: Average repeatability rate for each image sequence, scene and image format used. The RR is averaged over all frames in a sequence and over all four FP detectors tested. Significantly improved performances are typeset in bold.

When evaluating the viewpoint changing sequences, the middle images of the sequences were chosen as reference images, so RR decreases as the camera rotates both clockwise and counter-clockwise around the scene. In the case of the 3D scene, the RR is very similar regardless of image format used. But in the 2D scene, the LTM format provides very much higher RR (avg. 90 %) compared to the other formats (avg. 70 % – 71 %). This advantage applies to all four detectors tested, as can be seen in Figure 4(1b) and Table 1. The reason is that all features of the planar scene are visible from any direction. They gradually change their appearance because the brightness of the scene is changing with the rotation. When the camera reaches the position where it meets the reflected light of major light sources the brightness is maximal, but contrast decreases. The local tone mapper enhances local contrast and therefore makes the features look the same throughout the whole sequence. This is not the case with the other formats. This phenomenon does not occur in the case of the general (3D) scene since the light is reflected in many directions and, moreover, some features naturally disappear with the camera rotation. Closer examination of the diagrams in Figure 4(2c,2d) reveals that the FAST detector performed worse because it also detected random FPs on edges rather than just corners, as previously discussed.

When evaluating the lighting changing sequences, the images captured with all three light sources switched on were chosen as reference images. The experiments showed that the GTM format is most

¹Camera Calibration Toolbox for Matlab by Jean-Yves Bouguet: http://www.vision.caltech.edu/bouguetj/calib_doc/

²Automatic Camera Pose Estimation from Dot Pattern by George Vogiatzis and Carlo Hernández: <http://george-vogiatzis.org/calib/>

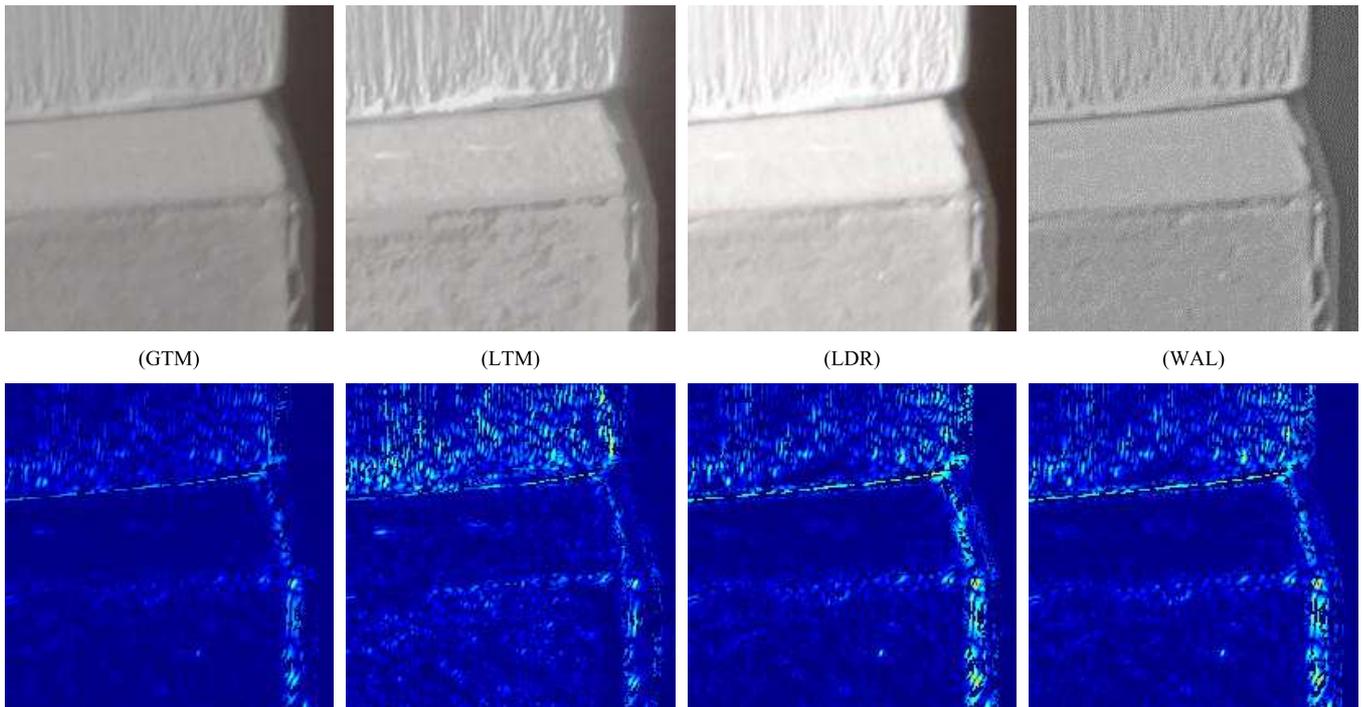


Figure 3: Different responses of the FAST detector for the same image region in the four image formats used. Setting the same response threshold for all image formats would result in significantly different numbers of detected FPs. Top row from left to right: Corners of 2 boxes from the 3D scene shown in GTM, LTM, LDR and WAL formats. Bottom row: The detector response in false colours for those image areas - from dark blue (no response) to red (highest response in this image area).

useful in cases of changing lighting conditions because it represents the scene and the lighting conditions most accurately. These conditions cause different parts of the scene to be under- or over-exposed in different images when using LDR imagery. Features in those parts disappear completely and cannot be detected, of course. This is not an issue when using HDR. The improvement of RR when using GTM instead of the LDR image format is from 64% to 67% in the case of 2D scene and from 40% to 55% in the case of 3D scene, respectively, as can be seen in Figure 4(3a,4a) and Table 1. Surprisingly, the LTM format can be disadvantageous if there are strong lighting gradients in the image, altering appearance of features under those gradients. This is the case of the image with light configuration 110 in Figure 4(3b). All image formats performed similarly in the 3D scene when the strip lights were on – see Figure 4(4a-4d), light configurations 100 – 111. Since the illumination generated by the strip lights has the character of an ambient light, it reduces the influence of the strong directional light generated by the reflectors used. The lighting conditions are therefore not so extreme and the scenes can be represented by all four image formats without any significant loss of detail.

When evaluating the distance changing sequences, the closest shots were chosen as reference images, so RR decreases as the camera moves away from the scene. If the distance between the camera and scene changes, the amount of light coming into the camera lens might change as well, causing the scene appearance to be altered (e.g. if the scene background is dark, moving the camera away from the scene will cause the scene to be over-exposed and features to be lost). The findings from the lighting changing scenarios are therefore generally valid for distance changing sequences as well. The scene over-exposure issue arose when capturing the 2D scene from distances 197 and 297 cm so we decided not to use these images in our evaluations since they would bias the results towards

the GTM format. After this correction, all image formats provided an average RR of 85% with exception of LTM (92%). In the case of the 3D scene, the GTM format proved to provide highest average RR 59%, contrary to other formats which provided an average RR of just around 49-52%, as shown in Table 1. The FAST detector performed distinctively worse again on nearly all formats, see Figure 4(6a,6c,6d). The only exception was the LTM format, see Figure 4(6b), where even fine details were distinct enough to be detected by FAST.

Any of the four detectors tested in this paper didn't outperform the others significantly on a single image format. We have shown that the Fast Hessian outperforms the other detectors slightly in most cases, while FAST usually performs slightly worse than others. Similar findings were published by Gauglitz et al. [2011]. In addition, Gil et al claimed that the Harris corner detector showed best results being closely followed by Fast Hessian detector and that these two behave in a similar way [Gil et al. 2010]. On the other hand, studies claiming that the FAST detector is the best one can also be found (e.g. [Jazayeri and Fraser 2010]), but these do not compare FAST with detectors which have been shown to outperform it, such as Harris corner detector, Shi-Tomasi or Fast Hessian, used in this paper.

5 Conclusions

In this paper we carried out a comparison of four different feature point detectors on four different image formats under extreme lighting conditions. These conditions were achieved by illuminating a planar and a 3D scene with reflectors in a totally dark room. The scenes were captured in three image sequences with varying cam-

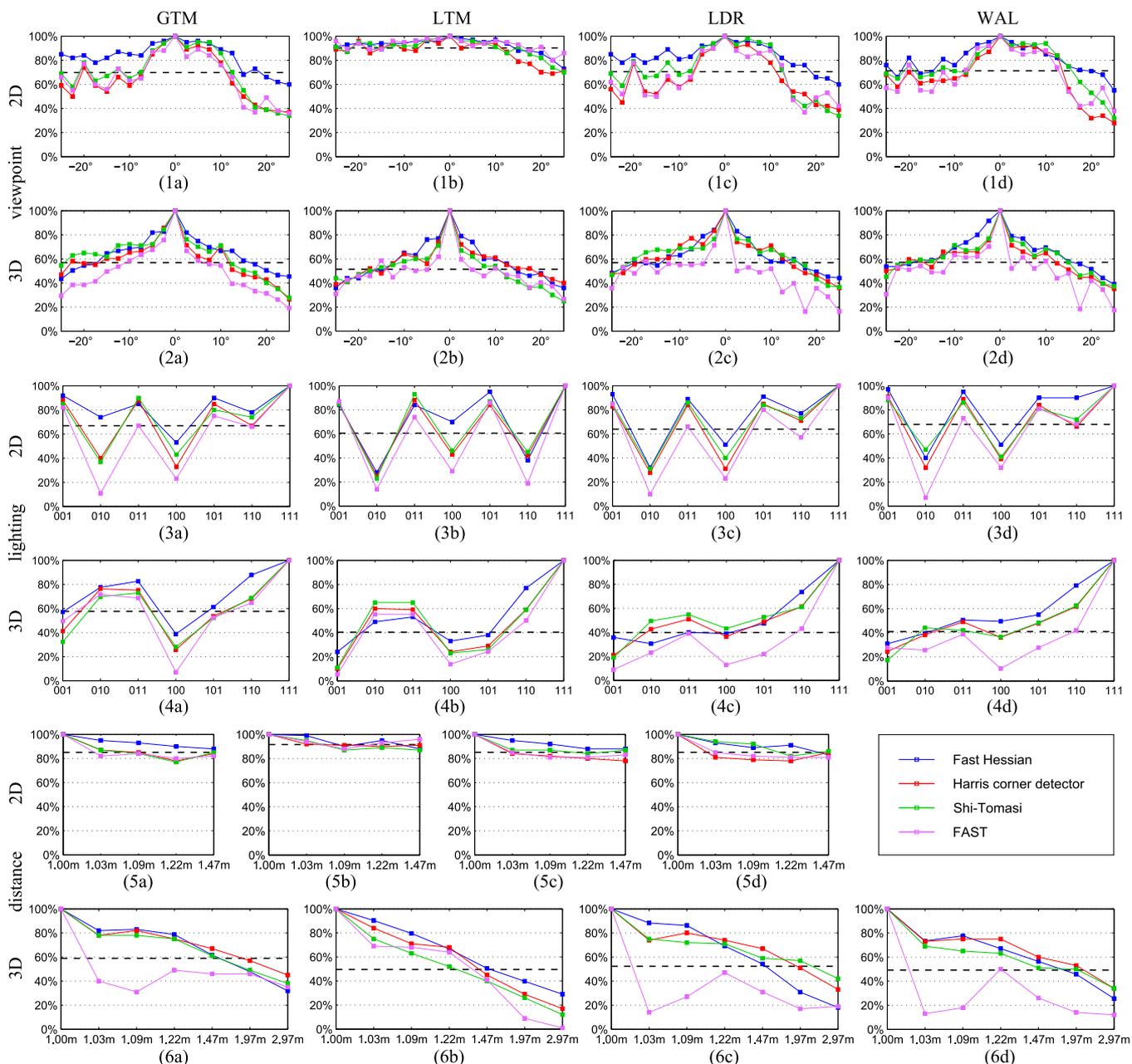


Figure 4: Repeatability rate of detectors under various conditions on various image formats. Rows 1 and 2 show results for **viewpoint** changing sequences, rows 3 and 4 for **lighting** changing sequences and rows 5 and 6 for **distance** changing sequences. Odd rows are for **2D** scenes, even rows for **3D** scenes. Each row consists of 4 diagrams showing results for different image formats - from left to right: **GTM**, **LTM**, **LDR** and **WAL**. Each diagram shows repeatability rate for each of the four tested FP detectors in each frame of the sequence. The horizontal dashed line in each diagram depicts the **average repeatability rate** of all detectors and frames for given image format.

era viewpoints, camera distances and scene lighting. The images of the scenes were stored into various image formats including tone mapped HDR images and ordinary LDR images. In particular, we investigated the repeatability rate of the detectors with respect to the image formats used.

Our results show that the locally tone mapped HDR format (LTM) enhanced the repeatability rate (RR) of the detectors from 71 % to 90 % in the case of the 2D scene, but only if there were no strong lighting gradients in the scene. The globally tone mapped HDR format (GTM) enhanced the RR when the lighting was changing,

especially in the case of the 3D scene (from 40 % to 55 %). HDR imaging techniques thus have an important role to play in improving the performance of FP detectors.

The improvement in performance of the detectors on Wallis-filtered images was quite marginal, so the Wallis filter is probably more suitable for pre-processing of remotely sensed images, where it has previously been shown to be successful [Gruen and Li 1995]. Such images typically have very low contrast and near-constant brightness, contrary to images we used in this work.

We observed that the Fast Hessian detector performed slightly better than others while the FAST detector performed slightly worse. This complies with previously published studies.

Future work will evaluate the influence of various tone mappers and their parameters on the performance of FP detectors. The simple global and local tone mapper used in this work showed to be beneficial in terms of the repeatability of the FP detectors, but there are many other tone mappers to be evaluated, both physically-based and perceptual-based [Ledda et al. 2005]. Future work will also implement the state-of-the-art detectors so they can work directly with the floating-point representation of HDR images. After this, an evaluation will be carried out into the evaluation of localisation accuracy of the detectors directly on HDR imagery, compared to the results obtained with HDR data after it has been tone mapped.

Acknowledgements

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic by project MSM 0021630528 “Security-Oriented Research in Information Technology”, by the Faculty of Information Technology, Brno University of Technology by project FIT-S-11-2 “Advanced Recognition and Presentation of Multimedia Data” and also partly by EU COST Action IC1005.

Our thanks also go to our WMG colleagues Elmedin Selmanovic, Jassim Happa, Thomas Bashford-Rogers and Samuel Staton for their comments and support.

The images depicted on posters used in the 2D scene were downloaded from the HD Wallpapers website³ and may be used for non-commercial purposes only.

References

- BANTERLE, F., ARTUSI, A., DEBATTISTA, K., AND CHALMERS, A. 2011. *Advanced High Dynamic Range Imaging: Theory and Practice*. AK Peters (CRC Press), Natick, MA, USA.
- BAY, H., TUYTELAARS, T., AND VAN GOOL, L. 2006. Surf: Speeded up robust features. *European Conference on Computer Vision 2006*, 404–417.
- CUI, Y., PAGANI, A., AND STRICKER, D. 2011. Robust point matching in hdri through estimation of illumination distribution. In *Proceedings of the 33rd international conference on Pattern recognition*, Springer-Verlag, 226–235.
- FRAUNDORFER, F., AND BISCHOF, H. 2005. A novel performance evaluation method of local detectors on non-planar scenes. In *IEEE Conference on Computer Vision and Pattern Recognition – Workshops*, IEEE.
- GAUGLITZ, S., HÖLLERER, T., AND TURK, M. 2011. Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision* 94, 3, 335–360.
- GIL, A., MOZOS, O., BALLESTA, M., AND REINOSO, O. 2010. A comparative evaluation of interest point detectors and local descriptors for visual slam. *Machine Vision and Applications* 21, 6, 905–920.
- GRUEN, A., AND LI, H. 1995. Road extraction from aerial and satellite images by dynamic programming. *ISPRS Journal of Photogrammetry and Remote Sensing* 50, 4, 11–20.
- HARRIS, C., AND STEPHENS, M. 1988. A combined corner and edge detector. In *Alvey Vision Conference*, vol. 15, Manchester, UK, 50.
- HARTLEY, R., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press.
- JAZAYERI, I., AND FRASER, C. 2008. Interest operators in close-range object reconstruction. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37, B5, 69–74.
- JAZAYERI, I., AND FRASER, C. 2010. Interest operators for feature-based matching in close range photogrammetry. *The Photogrammetric Record* 25, 129, 24–41.
- LEDDA, P., CHALMERS, A., TROSCIANKO, T., AND SEETZEN, H. 2005. Evaluation of tone mapping operators using a high dynamic range display. *ACM Transactions on Graphics* 24, 3, 640–648.
- MIKOLAJCZYK, K., TUYTELAARS, T., SCHMID, C., ZISSERMAN, A., MATAS, J., SCHAFFALITZKY, F., KADIR, T., AND GOOL, L. V. 2005. A comparison of affine region detectors. *International Journal of Computer Vision* 65, 1 (November), 43–72.
- MOREELS, P., AND PERONA, P. 2007. Evaluation of features detectors and descriptors based on 3d objects. *International Journal of Computer Vision* 73, 3, 263–284.
- OHDAKE, T., AND CHIKATSU, H. 2005. 3d modelling of high relief sculpture using image-based integrated measurement system. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36, 5/W17, 6.
- REMONDINO, F., AND ZHANG, L. 2006. Surface reconstruction algorithms for detailed close-range object modeling. In *Proceedings of ISPRS Commission III Symposium*, 117–123.
- RODEHORST, V., AND KOSCHAN, A. 2006. Comparison and evaluation of feature point detectors. In *5th International Symposium Turkish-German Joint Geodetic Days*.
- ROSTEN, E., AND DRUMMOND, T. 2005. Fusing points and lines for high performance tracking. In *IEEE International Conference on Computer Vision*, vol. 2, 1508–1511.
- SCHMID, C., MOHR, R., AND BAUCKHAGE, C. 2000. Evaluation of interest point detectors. *International Journal of computer vision* 37, 2, 151–172.
- SHI, J., AND TOMASI, C. 1994. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 593–600.
- WALLIS, K. 1974. Seasonal adjustment and relations between variables. *Journal of the American Statistical Association* 69, 345 (March), 18–31.
- ZHANG, Y., ZHU, Q., YU, J., AND ZHANG, Y. 2009. Automatic image mosaic-building algorithm for generating facade textures. *3D Geo-Information Sciences*, 257–269.
- ZUIDERVELD, K. 1994. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, Academic Press Professional, Inc., 474–485.

³<http://www.hdwallpapers.in>