J. Vis. Commun. Image R. 38 (2016) 141–160

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Evaluation of feature point detection in high dynamic range imagery ☆

Bronislav Přibyl [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

ARTICLE INFO

Article history:
Received 2 September 2014
Accepted 15 February 2016
Available online 24 February 2016

Keywords:
Feature point detection
Interest point detection
Corner point detection
Repeatability rate
Distribution of feature points
High dynamic range imagery
HDR
Tone mapping

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.

1. Introduction

Many computer vision tasks, such as image analysis, registration and indexing, object tracking, 3D reconstruction, and visual navigation (SLAM), 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/m^2 (on a starlit night) and 10^6 cd/m^2 (on a sunny day), see [3]. Such a difference between the luminance levels can generate a dynamic range of $1:10^9$, or 30 stops.[1] 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 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 [1,2,4–6] and application specific ones [7–10]. 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

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail addresses: ipribyl@fit.vutbr.cz (B. Přibyl), alan.chalmers@warwick.ac.uk (A. Chalmers), zemcik@fit.vutbr.cz (P. Zemčík), l.hooberman@warwick.ac.uk (L. Hooberman), cadik@fit.vutbr.cz (M. Čadík).

[1] The 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.

http://dx.doi.org/10.1016/j.jvcir.2016.02.007
1047-3203/

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 analyzed 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 autocorrelation function reflecting local intensity changes in the image [17]. For each point $\mathbf{x}$, the second moment matrix

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

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

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

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 $\mathbf{M}(\mathbf{x})$:

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

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

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

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

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

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

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

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 $3 \times 3 \times 3$-neighborhood 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 scene.[2] 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 fact.[3] 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

[2] The 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 meter ($\mathrm{cd/m^2}$) in calibrated images.

[3] Humans, 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 ($dL/L = \mathrm{const.}$). 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.