On Recall Rate of Interest Point Detectors

Henrik Aanæs
haa@imm.dtu.dk
DTU Informatics,
Technical University of Denmark

Anders Lindbjerg Dahl
abd@imm.dtu.dk
DTU Informatics,
Technical University of Denmark

Kim Steenstrup Pedersen
kimstp@diku.dk
Department of Computer Science,
University of Copenhagen, Denmark

Abstract

In this paper we provide a method for evaluating interest point detectors independently of image descriptors. This is possible because we have compiled a unique data set enabling us to determine if common interest points are found. The data contains 60 scenes of a wide range of object types, and for each scene we have 119 precisely located camera positions obtained from a camera mounted on an industrial robot arm. The scene surfaces have been scanned using structured light, providing precise 3D ground truth. We have investigated a number of the most popular interest point detectors. This is done in relation to the number of interest points, the recall rate as a function of camera position and light variation, and the sensitivity relative to model parameter change. The overall conclusion is that the Harris corner detector has a very high recall rate, but is sensitive to change in scale. The Hessian corners perform overall well followed by MSER (Maximally Stable Extremal Regions), whereas the FAST corner detector, IBR (Intensity Based Regions) and EBR (Edge Based Regions) performs poorly. Furthermore, the repeatability of the corner detectors is quite unaffected by the parameter setting, and only the number of interest points change.

1. Introduction

The ability to match descriptors obtained from local interest points is widely used for obtaining image correspondence. This is based on the assumption that it is possible to find common interest points. For this to be useful for e.g. geometric reconstruction, corresponding interest points have to be localized precisely on the same scene element, and the associated region around each interest point should cover the same part of the scene. In general it is hard to determine if correspondence exist between interest points, because it requires ground truth of the geometry of the observed scene.

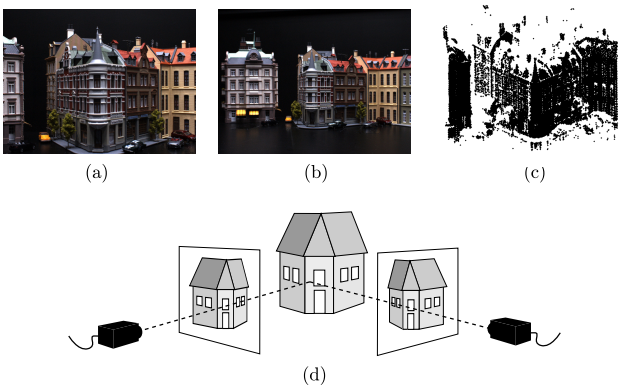

Figure 1. Example of data and setup. Two images of the same scene with one close up (a), one distant from the side (b), and the reconstructed 3D points (c). Corresponding interest points can be found using the geometric information of the scene with known camera positions and 3D scene surface as schematically illustrated in (d).

A wide range of applications are based on matching local image descriptors obtained from image interest points. These applications include object recognition [8], image retrieval [15, 20], and similar, but for these types of applications the precision of the spatial position is less important. However in applications for 3D geometry reconstruction from interest points it is important to have a precise point correspondence [21, 22, 24].

It is common to distinguish between detecting interest points and computing the associated descriptor. This could indicate that the two steps are independent, see e.g. [11, 12].

The question is, however, if this assumption is reasonable. Interest points and the associated regions are found from salient image features, and the same image features will be part of the actual characterization. As a result the two parts are not independent, and the choice of interest point detector will influence the description of the region around the interest point. This will limit the subspace spanned by the descriptors and this way reduce the specificity of the descriptor. An alternative to feature based interest points is to pick the interest points at random, but it will be unlikely to obtain precise spatial correspondence between a sparse set of randomly picked point. So, the ability to detect corresponding interest points, in a precise and repeatable manner, is a desirable property for obtaining geometric scene structure. In this paper we will investigate exactly that property.

Early work on correspondence from local image features was based on rotation and scale invariant characterization [8, 18], and interest points from planer scenes was evaluated in [19]. Later the interest points has been adapted to affine transformation, making the characterization robust to larger viewpoint change. These methods have been surveyed in [12], but the performance has been evaluated on quite limited data-sets, consisting of ten scenes each containing six images. The suggested evaluation criteria have since been used in numerous works together with this small data set.

The ground truth in the data from [12] was obtained by semi-manually fitting an image homography. As a consequence this limits the scene geometry to planar surfaces or images from a large distance where a homography is a good approximation. To address this issue Fraundorfer and Bishof [4] generated ground truth by requiring that a matched feature should be consistent with the camera geometry across three views. In Winder *et al*. [7, 27, 28] results from Photo Tourism [21] was used as ground truth. These approaches use feature matching to evaluate the matching of features, which can be problematic. If errors occur in the ground truth there can be a bias towards wrong correspondences in the proposed matching. As a result these wrong correspondences will not be detected.

Moreels and Perona [13] evaluated interest point features similar to [4] based on pure geometry by requiring three view geometric consistency with the epipolar geometry. In addition to this they used a depth constraint based on knowledge about the position of their experimental setup. Hereby they obtained unique correspondence between 500-1000 interest points they obtain from each object. The limitation of their experiment is relatively simple scenes with mostly single objects resulting in little self-occlusion, which is very frequent in real world scenes and typically many interest points are found near occluding boundaries.

We have compiled a large data set that provides a unique basis for this study. It consists of 60 scenes of varying object types and complex surface structures resulting in a total of almost 150.000 images[1]. Figure 1 shows an example from our data set. The experimental setup consists of a camera mounted on an industrial 5-axis robot-arm, providing accurate and repeatable positioning. In addition the scenes have been surface scanned using structured light, and together with the camera positions these scans supply ground truth for feature correspondence. As a result we can easily find corresponding interest points on the scene surface. We evaluate nine established interest point detectors on this data set and provide new insight into the stability of these detectors with respect to large viewpoint and scale change.

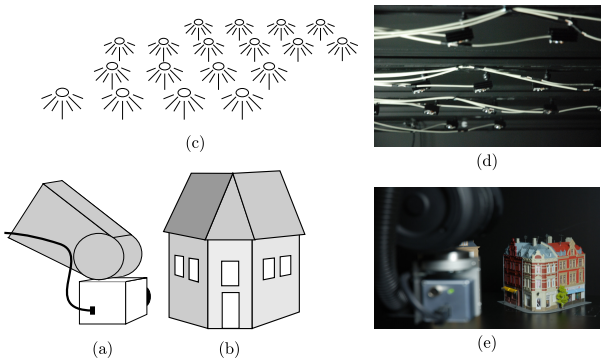

Figure 2. Illustration of data collection setup. The camera is mounted on a robot arm (a) capturing images of the scene (b). LED point light sources illuminate the scene from 18 individual positions (c). Photos of the real setup (d,e).

The paper is organized as follows. Section 2 describes the image data and details of how we obtained ground truth. In Section 3 our method for geometric based correspondence is described, and in Section 4 the details of the experiments are given. We discuss and conclude the approach and obtained results in Sections 5.

2. Data

The setup for data acquisition is illustrated in Figure 2, and a detailed description of the data is available in [1]. The entire setup is enclosed in a black box and the scenes can be up to about half a meter, but the closest images depict about 25×35 cm. Scenes have been selected to show a large variation in scene type and they contain elements that are challenging for computer vision methods, like occlusions and various surface reflectance properties. There are 60 scenes with varying type of material and reflectance properties, including model houses, fabric, fruits and vegetables, printed media, wood branches, building material, and art objects. Image examples are shown in Figure 4. Color images of 1200×1600 pixels have been acquired, but we use 600×800 down-sampled versions in grayscale, which is done for computational reasons.

[1] http://www.imm.dtu.dk/robotData

Camera positions The images are acquired from a camera mounted on a robot arm. This gives a freedom to position the camera in any direction. Figure 3 illustrates the camera path that we used for each scene. We have chosen the center frame closest to the scene as a *key frame*. The key frame is used as the reference image for evaluation, and in our experiments feature correspondence is found relative to that frame. The camera positions are very accurate with a standard deviation of approximately 0.1 mm, and the standard deviation on the back projected pixel was 0.2 to 0.3 pixels.

Light The scene is illuminated by 18 individually controlled light emitting diodes (LEDs), which can be combined to provide a highly controlled and flexible light setting, see e.g. [6]. We use this in an experiment with varying illumination, and in the other experiments we have simple combination of all light sources for diffuse illumination.

Surface reconstruction We use structured light to obtain 3D surface geometry of the scenes. Figure 1 (c) shows an example of our surface reconstruction. We use a stereo camera setup and gray encodings to solve the correspondence problem. This method is recommended as one of the most reliable methods in both Scharstein and Pal [17] and Salvie *et al*. [16]. The scene surfaces have been scanned using two camera pairs at two distances from the scene. This is done to cover as much of the scene visible from the key frame as possible, see Figure 3. We obtain a varying number of surface points from around 100.000 to 500.000 points depending on the size of the scene.

3. Method

Our goal is to analyze invariance properties of interest points found in corresponding images. The design of our data set enables us to answer questions like what happens to interest points with change in view point? How many of the interest points are actually relevant? Are interest points precisely located? Answers to these and related questions will provide an improved basis for choosing the appropriate methods for extracting interest point for computer vision system design. Evidence for interest point correspondence is obtained by fulfilling three criteria. We will now provide the details of our analysis.

The normal procedure for image matching using interest point descriptors contains the following three steps. First interest points are detected providing a spatial localization of the regions of interest. Secondly a descriptor is assigned to the detected interest point, which is invariant to scale, rotation, and affine transformation. In the third step the descriptors are matched to find correspondence.

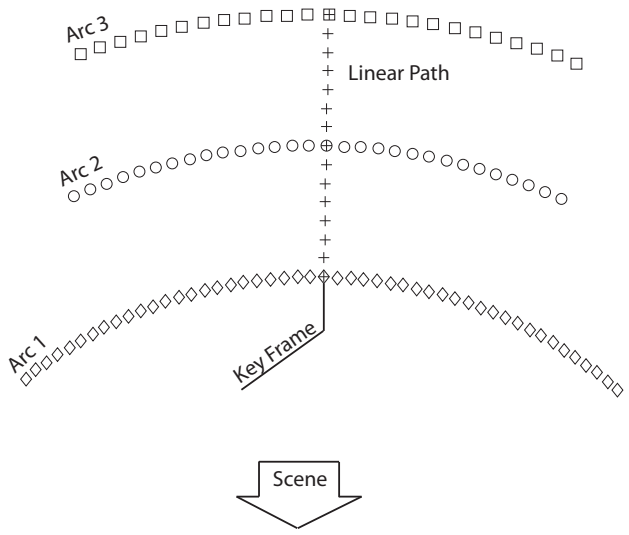

Figure 3. Camera positions. The camera is placed in 119 positions in three horizontal arcs and a linear path away from the scene. The central frame in the nearest arc is the key frame, and the surface reconstruction is attempted to cover most of this frame. The three arcs are located on circular paths with radii of 0.5 m, 0.65 m and 0.8 m, which also defines the range of the linear path. Furthermore, Arc1 spans $+/-40^\circ$, Arc2 $+/-25^\circ$ and Arc3 $+/-20^\circ$.

Evaluation criteria Evaluation of the performance of interest point detectors cannot be based on the associated descriptor, because the descriptors might not be unique. As a result it is impossible to tell if a given correspondence is correct or a mismatch between similar looking image regions. Therefore the evaluation has to be done independently of the interest point detection. We utilize the geometry of both the 3D scene surface and the camera positions to obtain this independent evaluation basis. Our evaluation criteria, with regard to pixel distances and scale, are based on a trade-off between as few double matches as possible and not discharging points because of small variations in position of the interest points.

For each point in the key frame there has to be at least one interest point in the corresponding image fulfilling all three criteria, for the point to count as having a potential match. If more than one point fulfill all criteria it still counts as a one potential match.

Epipolar geometry Consistency with epipolar geometry is the first evaluation criterion. The camera positions of all images in our data set is known with high precision, which provides basis for the relationship between points in one image and associated epipolar lines in another. This is used for removing false matches for a given interest point. We discharge points that are further away than 2.5 pixels orthogonal to the epipolar line, as illustrated in Figure 5 (a).

Figure 4. Example images from our data set. The images show a diffuse relighting obtained by a linear combination of the 18 directional illuminated images. From left the scenes are examples of houses, books, fabric, greens, and beer cans, which has been used in our feature matching experiment with light variation.

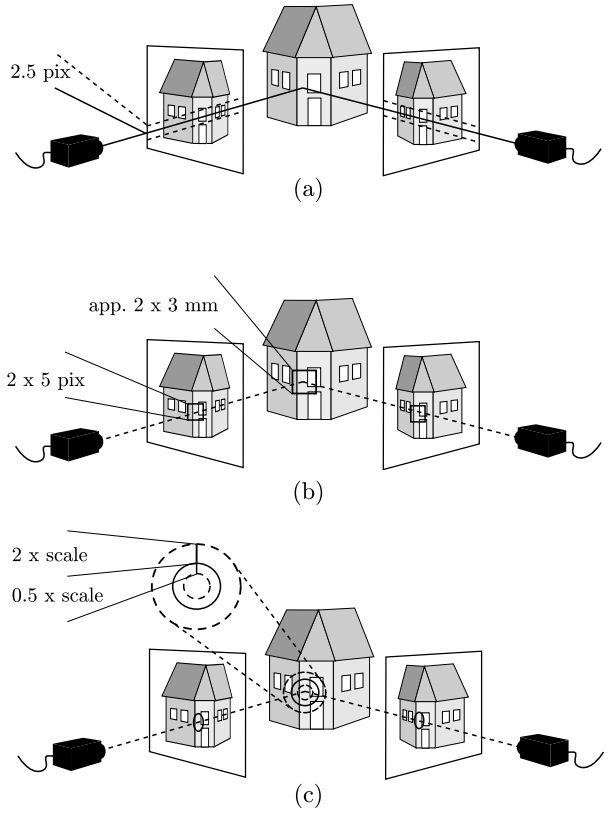

Figure 5. Matching criteria for interest points. This figure gives a schematic illustration of a scene of a house and two images of the scene from two viewpoints. (a) The consistency with epipolar geometry, where corresponding descriptors should be within 2.5 pixels from the epipolar line. (b) Window of interest with a radius of 5 pixels and corresponding descriptors should be within this window, which is approximately 3 mm on the scene surface. Ground truth is obtained from the surface geometry. (c) The scale consistency, where corresponding descriptors are within a scale range factor of 2 from each other.

Surface geometry 3D surface reconstruction is used in the second evaluation criterion. Two points are considered a positive match if their 3D position is close to the scene surface obtained from the structured light reconstruction. This is fulfilled if there is a point from the surface reconstruction within a window of 10 pixels around a point, which corresponds to a box of approximately 6 mm on the scene surface. The surface reconstruction is not complete, so points in regions without surface reconstruction are discharged, which only few points were. The surface geometry constraint is illustrated in Figure 5 (b).

Absolute scale A region around each interest point provides the basis for an image descriptor. The interest points are detected in a multi-scale approach and the size of this region is dependent on what scale the interest point is detected. This image region corresponds to an area on the scene surface and corresponding descriptors should cover the same scene part. This area correspondence provides the basis for the third evaluation criterion, which is illustrated in Figure 5 (c), and the area of this region has to be within an area range of 0.5 - 2 of each other.

4. Experiments

We have investigated a number of interest point detectors namely Harris and Hessian corner detectors [10], Maximally Stable Extremal Regions (MSER) [9], Intensity Based Regions (IBR) and Edge Based Regions (EBR) [26], and the Fast Corner Detector [25], and our experiments are based on the implementations presented in [11, 12][2]. As an evaluation measure we use recall rate, similar to that of [11], which is the ratio

$$\text{Recall} = \frac{\text{Potential Matches}}{\text{Total Interest Points}} \quad .$$

The potential matches are points from the key frame fulfilling all three correspondence criteria. The total number of interest points is the number of interest points found in the key frame, see Figure 3.

Methods for interest point detection should ideally identify the same scene regions independently of camera position and illumination. As a result we have investigated the recall rate of the interest point detectors relative to variation in camera position and light over the 60 scenes in our

[2] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html

Detector	# Interest Points	Std. Interest Points
Harris	925	665
Harris Laplace	736	538
Harris Affine	718	524
Hessian Laplace	1045	635
Hessian Affine	839	560
MSER	354	261
EBR	423	614
IBR	250	139
FAST	1539	1644

Table 1. Average number of interest points detected and the standard deviation over the 60 scenes.

data set. Furthermore, we have varied the input parameters for the methods to test if the algorithms are sensitive to parameter variation. First we will look at the detected number of interest points with the recommended parameter settings according to [9, 10, 11, 12, 25, 26].

Number of interest points A varying number of interest points are detected in each data set, but this is highly dependent on the detection algorithm and the depicted scene. Table 1 shows the number of interest points and the standard deviation relative to the 60 scenes, where interest points have been extracted with the recommended parameter values. Some variation in number of interest points will be expected, because of scene variation, but there is a noteworthy difference between the methods.

Especially the FAST corner detector has some scenes with nearly 10.000 interest points and other scenes with close to 0. This is especially undesirable since it appears that scenes exist for which this algorithm will not work. The EBR also has a large variation, but much fewer interest points, and the IBR has few interest points. Few interest points is an undesirable property because it makes it hard to estimate the image correspondence. But also large fluctuations will result in unpredictable running time, and especially a very large number of interest points can slow down the matching procedure. The Harris and Hessian corner detectors gives a reasonable number and variation of interest points, whereas MSER has relatively few points, but with a reasonable number in all scenes.

Recall and position The recall rate of the interest point detectors as a function of the camera position is shown in Figure 6. Interest point detectors are sensitive to the camera position, and both changing the angle to the scene and the distance will reduce the recall. The question is what shape we can expect the curves to have.

The statistics of objects in ensembles of natural scene exhibits statistical scale invariance [23]. Among other, this arises from the fact that the empirical distribution of area of homogeneous image segments follows a power law [2]. A recent study [5] shows that averaged over ensembles of scenes this area distribution appears to be invariant to change of distance to the scene. Even though our data set consists of indoor still leben scenes, we expect it to follow a power law behavior, especially because our scenes include the so-called "blue-sky effect" [14] in the form of the large black background area. Therefore, from these empirical results we may deduce that as the camera moves away from the scene, small details, including potential interest points at low scales, will disappear in large numbers, and few new large scale structures will appear leading to new potential large scale interest points. Since the distribution of structure follows a power law the consequence is that the number of interest points is expected to decrease as the viewing distance increases. This will in turn lead to a decrease of the recall rate. Hence for well-behaving interest point detectors we expect the recall rate to follow a decreasing power law as a function of viewing distance. Furthermore, we have no reason to prefer certain view orientations, hence we expect at least symmetry, if not rotational invariance, in the recall rate with respect to varying orientation.

The shape of the curves in Figure 6 behaves mostly as expected, i.e. following a power law. But the Harris corner detector performs very well at moderate scale change, but has a sudden drop in recall rate at a distance of 0.7 m (Figure 6 (d)). This indicates that a lag of scale invariance for this detector, which could be a matter of the implementation choices. To validate this further investigations are needed. Especially the Hessian corner detectors perform overall well, but also the Harris Affine and Harris Laplace corner detectors have good performance. The FAST corner detector also has a high recall rate, but exhibits asymmetries with respect to orientation, which might be caused by the large variation in the detected interest points. This can probably be explained by the large variation in the number of interest points detected in the various scenes.

Changing light The 18 LED's used in the data set constitutes a light stage [3]. From this setup the scenes can be synthetically illuminated by a linear combination of the images. The synthetic illumination was used to investigate the recall rate of the interest point detectors as a function of changing light.

Ten different illumination settings were constructed, varying the incident angle of the light, and furthermore the light was changed from diffuse to directional. The recall rate was computed between the tenth illumination with all the illuminations at a camera position[3] separated by 10°. Results are shown in Figure 7, where the observations are ordered according to common LED's used for the illumination. This proved to be the determining factor for the

[3]The results are similar for other positions.

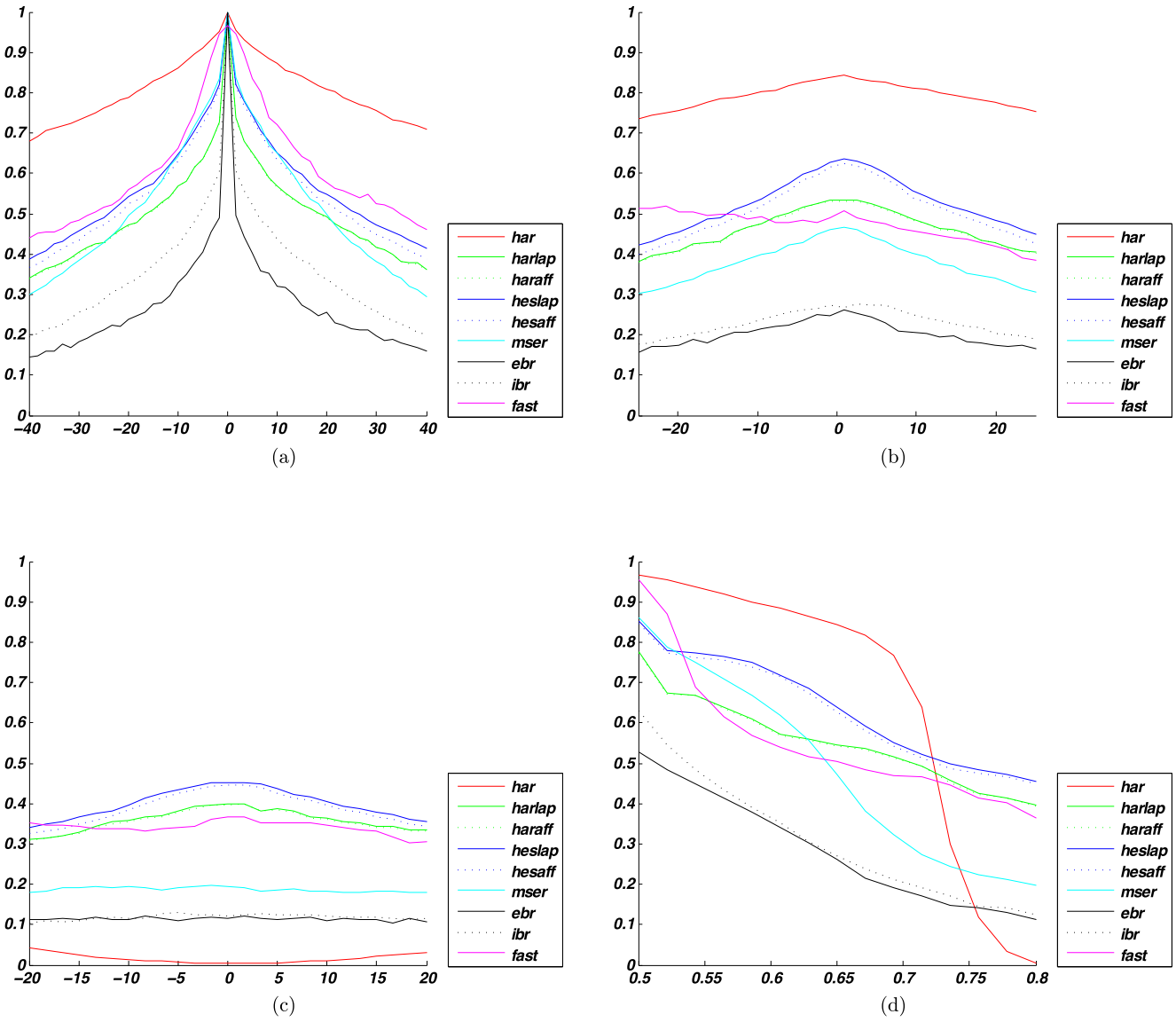

Figure 6. Mean recall rate. The graphs show the recall rate relative to the paths shown in Figure 3 with Arc 1 (a), Arc 2 (b), Arc 3 (c), and Line Path (d). The horizontal axis is the angle relative to the scene in (a-c) and distance to the scene in (d). The vertical axis is the recall rate.

performance of the detectors in terms of recall rate. The observed change in recall rate is expected because scenes change appearance with change in light, e.g. cause by shadows and other reflectance effects. The least sensitivity to light change with regard to recall rate is seen for the Harris corner detector followed by the Hessian corner detectors. Harris affine, Harris Laplace and MSER perform equally well, whereas EBR, IBR and FAST have a poor performance.

Changing model parameters In the above experiments the recommended parameter settings were used. These correspond to standard settings of the downloaded software. To investigate the effect of these settings, we conducted the experiments with varying position with different cornerness setting, for the Harris and Hessian type detectors. The parameter was varied on a logarithmic scale from $0.107 - 9.31\times$ the recommended parameter settings. This is done in 21 steps by a multiplicative factor of 1.25.

From these experiments we observe that the recall rate of the Harris type detectors are unaffected by a change in the cornerness parameters and that the Hessian type detectors are only moderately affected. This is despite the corenerness parameter drastically affecting the number of interest

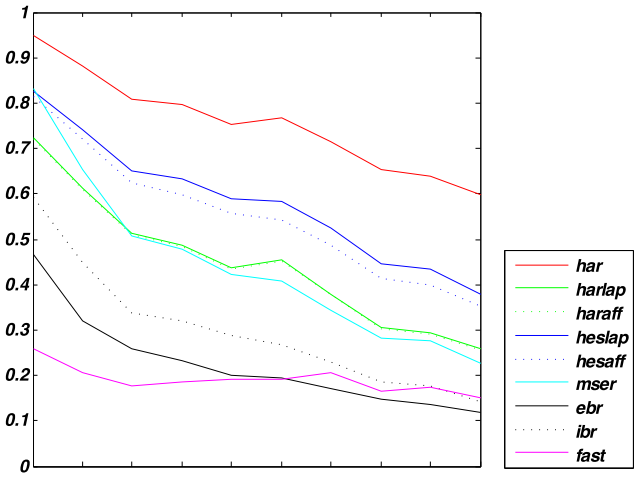

Figure 7. Mean recall rate relative to change in light direction for a camera position 10° from the key frame, averaged over all 60 scenes. The horizontal axis indicates the amount of common LED's being turned on, compared to the lighting used in the key frame.

Detector	Recall Rate Correlation
Harris	-0.0198
Harris Laplace	-0.0286
Harris Affine	-0.0275
Hessian Laplace	-0.2149
Hessian Affine	-0.1656

Table 2. The average correlation of the recall rage by changing the threshold parameter from $0.107 - 9.31\times$ the recommended parameter settings.

points extracted. These observations are quantisized in Table 2, which shows the correlation between the recall rate and the cornerness parameter setting. This implies that our results are relatively insensitive to the choice of parameter setting.

5. Discussion and Conclusion

The contribution of this paper is an investigation of nine established interest point detectors, which provides new insights to the stability of these detectors with respect to large changes in viewpoint and scale. The investigation is based on a data set of 60 scenes with precise ground truth of camera position and scene surface, acquired with an industrial robot arm. Furthermore, a controlled light setting has enabled us to perform precisely controlled relighting experiments.

This unique data set enables us to investigate interest point correspondence independently of descriptors for very complex, non-planar scenes. The key element that we investigate is if there for a given interest point is a potential matching interest point in a corresponding image. Our investigation is based on the same implementations as in the extensive study of interest points by Mikolajczyk and Schmid [12]. The novelty of our investigation is the complexity of the data set, both with regard to number of scenes and geometric surface structure.

The first investigation is the number of interest points provided by the algorithms. Most of the algorithms provide a reasonable number of interest points, but not the FAST corner detector, which is highly unstable in the number of interest points, ranging from close to 0 to around 10.000. This is very undesirable, because this method is unreliable for solving the correspondence problem. A similar property is seen for the EBR algorithm, but with fewer interest points, and the IBR has very few points. The best performance is the Harris and Hessian corner detectors, and MSER is also reasonably stable, but with relatively few interest points.

Secondly we have investigated the recall rate relative to camera position, which provides very interesting results. We expect the recall rate to follow a decreasing power law distribution as a function of viewing distance. This is also seen for most interest point detectors, but for most viewpoint changes the FAST corner detector does not show this behavior, see Figure 6, probably due to the high variation in number of detected interest points. The Harris corner detector performs very well for small-scale changes, but has a large drop in performance with scale change. Overall the Hessian performs slightly better than the Harris corner detector, and the corner detectors performs better than MSER. IBR and EBR perform poorly.

We have made an experiment with change in light from diffuse to directional. Again we obtain the best performance with the Harris corner detector followed by Hessian corners. Harris Laplace, Harris Affine and MSER perform equally whereas EBR, IBR and FAST performs poorly.

In [12] the parameter of the interest point detectors, which affects the number of detected interest points, is discussed. The claim is that a parameter choice favoring the most stable regions will give a high repeatability score, and with many detected interest points, clutter will similarly give a high repeatability. Their measure of repeatability is similar to the recall rate presented here. We have investigated parameter choice for Harris and Hessian corners, and our investigation contradicts this claim, given a reasonable choice of parameters. Especially the Harris corners are independent of the parameter choice whereas there is a higher correlation for the Hessian corners with higher recall for the most stable interest points, see Table 2.

Overall the simple Hessian corner detector performs very well, but is not invariant to scale change. Compared to the study in [12] we also get an overall good performance of the Hessian corner detectors. We have not seen as good a

performance of the MSER as is reported in that study, which might be caused by the non-planer scenes in our study.

Viewed from a pure interest point detection perspective, corner detectors perform better than region detectors and especially the EBR, IBR and FAST performs poorly. It is important to note, that this study only concerns interest points, which is just one element of solving the correspondence problem, and the success of a system will depend on the interest point descriptor and the matching procedure as well. But the insights brought by this study show a clear performance difference and what the effect of the interest point detector will be in a final system.

6. Acknowledgements

We would like to thank the Oxford Vision Group[4] for making their code available online. Furthermore, we would like to thank the Center for Imaging Food Quality, The Danish Research Council for financial support.

References

[1] Technical report. 2

[2] L. Alvarez, Y. Gousseau, and J.-M. Morel. The size of objects in natural and artificial images. In P. W. Hawkes, editor, *Advances in Imaging and Electron Physics*. Academic Press, 1999. 5

[3] P. Einarsson, C. Chabert, A. Jones, W. Ma, B. Lamond, T. Hawkins, M. Bolas, S. Sylwan, and P. Debevec. Relighting human locomotion with flowed reflectance fields. *Rendering Techniques*, pages 183–194, 2006. 5

[4] F. Fraundorfer and H. Bischof. Evaluation of local detectors on non-planar scenes. In *In Proc. 28th workshop of AAPR*, pages 125–132, 2004. 2

[5] D. Gustavsson. *On Texture and Geometry in Image Analysis*. PhD thesis, Department of Computer Science, University of Copenhagen, Denmark, June 2009. 5

[6] P. Haeberli. Synthetic lighting for photography. *Grafica Obscura*, 1992. 3

[7] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. *IEEE 11th ICCV*, pages 1–8, 2007. 2

[8] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004. 1, 2

[9] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004. 4, 5

[10] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *Int. J. of Computer Vision*, 60(1):63–86, 2004. 4, 5

[11] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE PAMI*, 27(10):1615–1630, 2005. 1, 4, 5

[4] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html

[12] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, 2005. 1, 2, 4, 5, 7

[13] P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. *Int. J. Comput. Vision*, 73(3):263–284, 2007. 2

[14] D. Mumford and B. Gidas. Stochastic models for generic images. *Quarterly of Applied Mathematics*, 59(1):85–111, March 2001. 5

[15] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *CVPR*, volume 5. Citeseer, 2006. 1

[16] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004. 3

[17] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *CVPR*, volume 1, pages 195–202, 2003. 3

[18] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997. 2

[19] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(4):151–172, 2000. 2

[20] J. Sivic and A. Zisserman. Video Google: Efficient visual search of videos. *Lecture Notes in Computer Science*, 4170:127, 2006. 1

[21] N. Snavely, S. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008. 1, 2

[22] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, 80(2):189–210, November 2008. 1

[23] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. On advances in statistical modeling of natural images. *Journal of Mathematical Imaging and Vision*, 18(1):17–33, January 2003. 5

[24] P. Torr and A. Zisserman. Feature based methods for structure and motion estimation. *Lecture notes in computer science*, pages 278–294, 1999. 1

[25] M. Trajković and M. Hedley. Fast corner detection. *Image and Vision Computing*, 16(2):75–87, 1998. 4, 5

[26] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004. 4, 5

[27] S. Winder, G. Hua, and M. Brown. Picking the best daisy. In *CVPR*, Miami, June 2009. 2

[28] S. A. J. Winder and M. Brown. Learning local image descriptors. In *CVPR*, pages 1–8, 2007. 2