

Point-of-Interest Detection for Range Data

Fredrik Viksten
Div. of Information Coding

Klas Nordberg
Computer Vision Lab.

Mikael Kalms
Div. of Information Coding

Linköping University, Sweden
{viksten,klas}@isy.liu.se / mikael@kalms.org

Abstract

Point-of-interest detection is a way of reducing the amount of data that needs to be processed in a certain application and is widely used in 2D image analysis. In 2D image analysis, point-of-interest detection is usually related to extraction of local descriptors for object recognition, classification, registration or pose estimation. In analysis of range data however, some local descriptors have been published in the last decade or so, but most of them do not mention any kind of point-of-interest detection. We here show how to use an extended Harris detector on range data and discuss variants of the Harris measure. All described variants of the Harris detector for 3D should also be usable in medical image analysis, but we focus on the range data case. We do present a performance evaluation of the described variants of the Harris detector on range data.

1. Introduction

The detection of robust points-of-interest, POIs, in a 2D image is important for many applications such as image registration, tracking and stereo matching, the latter was one of the first applications using local POIs [17]. Almost a decade later the Harris detector [9] was presented and subsequently used for much the same application. Matching Harris points over large distances is somewhat problematic and in [27] it is shown that this process can be made more robust by using the surrounding neighborhood around each POI to select which matches were likely. The use of local neighborhoods in [27] did not use any invariants at all as descriptor, something that was shown to be of great use in [20]. This has since then been followed by a range of POI-detectors and local descriptors [14, 2, 16, 15], which have been at the base of many successful applications.

In the area of POI-detection and local descriptors

for range data, most publications have focused on the later part. Publications like [4, 12, 11, 5, 8] mention no POI-detection at all, [21] would probably not be regarded as using a POI-detector and [22, 24] do detect POIs, but only after the extraction of local descriptors. In [25, 3, 7] POI-detectors are used, but none of them use a version of the Harris detector. In medical image analysis some POI-detectors have been developed. In [23], extremal points of the two main curvature directions are used as POIs. A comparison of 9 different differential operators [10] reaches the conclusion that POI-detectors based on the Harris matrix performs better than the ones based on measures of curvature. The measures on the Harris matrix used in [19, 1, 18, 10] are mostly concerned with the determinant, the trace and fractions between them. Given the lack of POI-detectors for range data, the superiority of differential-based detectors over curvature-based detectors for medical data and that most available range data detectors are based on curvature we will here focus on getting variants of the Harris detector working for range data and evaluate its performance.

The Harris detector has in [13] been extended to three dimensions, two spatial dimensions and time, which is the only publication among those mentioned here that explicitly mentions its work as an extension to the Harris detector. The formula used is very similar to the original Harris detector but uses only two out of three possible invariants under rotation of the Harris matrix in 3D. We will here extend the Harris measurement to include all three possible invariants under rotation of the Harris matrix and show that this can benefit the POI detection in range data.

2. Range Data Pre-processing

Range data is different from other 3D data in that it has a grid in only two, or fewer, directions. Our range data has a grid in two dimensions but not in the third, but

our proposed techniques should easily extend to other range data as well. In this work we make use of conservative surface voxelization [6, 26], where triangles are scan-converted within a volume between all neighbouring points in the range data 2D grid. This means that we will add data points where we in fact did not measure anything. The voxelization gives us a full 3D volume representation of the surfaces we have measured with our range data camera. One argument against the voxelization could be that it will yield a big explosion in the amount of data, but in practise it is not, since most of the volume is empty. The resulting volume is heavily sparse; a typical range scan occupies only one percent of the full volume. Most real range data, except for data captured by time of flight devices, have areas where self-occlusion leads to missing data. If there are points of missing data in the 2D grid, those points are not connected in our voxelization process, neither are points on either side of the missing data. In [7] a full volume was also generated, but it instead focused on the density of range data points as a measure of certainty. We use the model of straight hyper-lines between range data points and assume that sparsity of range points is an artifact of slanting surfaces which should not be interpreted as certainty. Many range sensors do have a separate certainty measure which could be used instead. One might argue that voxelization distorts the representation of the object so that it no longer looks like the real object. In object detection, recognition, pose estimation etc. we argue that it does not matter that we change the appearance of the object as long as it is changed the same way every time we use the data. We do however argue that the voxelization process has a strong benefit; it enables us to use differential-based detectors such as the Harris detector on the data.

3. The Harris Detector

The Harris detector tries to find positions in an image where the intrinsic dimension is larger than 1, i.e. where the signal varies in more than one direction. This is done by analysis of the eigenvalues, λ_i , of the Harris matrix, which for a 2D image f is given by

$$H_{2D} = \text{local weighted mean of } \begin{pmatrix} f_x^2 & f_x f_y \\ f_x f_y & f_y^2 \end{pmatrix} \quad (1)$$

where subscript of f signify derivative in the given direction. The analysis of eigenvalues is done by the Harris measure which is

$$m(k) = \det(H_{2D}) - k \text{trace}(H_{2D}) \quad (2)$$

where k is a parameter. For positions where the intrinsic dimension is 2 the eigenvalues should lie on a line

defined as $l = (\lambda_1 \lambda_2)^T = \lambda(1 \ 1)^T$, i.e. in 45 degrees angle from the origin in λ -space. The parameter k enables non-isotropic signal regions to be detected, i.e. positions in λ -space close to the line l but not exactly on it. This can be seen by looking at iso-contours for $m(k)$ which can be bent away from the two axes in λ -space by changing k .

Generalizing the Harris detector to 3D is based on the 3D Harris matrix

$$H_{3D} = \text{l.w.m. of } \begin{pmatrix} f_x^2 & f_x f_y & f_x f_z \\ f_x f_y & f_y^2 & f_y f_z \\ f_x f_z & f_y f_z & f_z^2 \end{pmatrix} \quad (3)$$

and can be done in some different ways. In [19, 1, 18, 13], variations of the measure $m(\cdot)$ use the determinant and trace

$$\det(H_{3D}) = \lambda_1 \lambda_2 \lambda_3 \quad (4)$$

$$\text{trace}(H_{3D}) = \lambda_1 + \lambda_2 + \lambda_3 \quad (5)$$

of the Harris matrix. The determinant and the trace of the matrix are used because they are invariant under rotation of the underlying signal, but none of the mentioned references make use of the third invariant, given by

$$\text{sec}(H_{3D}) = \lambda_1 \lambda_2 + \lambda_2 \lambda_3 + \lambda_3 \lambda_1. \quad (6)$$

We refer to it as $\text{sec}()$ since it is the second order invariant. Similarly to the 2D case, the spatial positions of intrinsic signal dimension 3 should lie on the line $l = (\lambda_1 \lambda_2 \lambda_3)^T = \lambda(1 \ 1 \ 1)^T$ in eigenvalue-space. The inclusion of the second order term gives us the opportunity to form other kinds of Harris extensions than what is found in [13, 10]. We will evaluate the following expressions:

$$D1 = \det(H) - l \text{sec}^{3/2}(H) \quad (7)$$

$$D2 = \det(H) - l \text{sec}(H) \text{tr}(H) \quad (8)$$

$$D3 = \det(H) - k \text{tr}^3(H) - l \text{sec}^{3/2}(H) \quad (9)$$

$$D4 = \det(H) - k \text{tr}^3(H) - l \text{sec}(H) \text{tr}(H) \quad (10)$$

and compare to these from the literature

$$Op3 = \frac{\det(H)}{\text{tr}(H)} \quad , \text{ see [10]} \quad (11)$$

$$Rohr3D = \det(H) \quad , \text{ see [10]} \quad (12)$$

$$Laptev = \det(H) - k \text{tr}^3(H) \quad , \text{ see [13]} \quad (13)$$

Having two parameters gives us more control over the form of non-isotropic signals we should detect. The $\text{sec}()$ term can be used to bend away an iso-contour from the planes spanned by any two of the base directions in λ -space.

The following parameters for D1-D4, see table 1, were found through a grid search over k and l where detected corners were compared to handmarked corners.

	D1	D2	D3	D4
k	-	-	0.0015	0.001
l	0.014	0.007	0.004	0.004

Table 1. Parameters for D1-D4.

4. Performance Evaluation

We used two sets of evaluation data in our performance comparison; first we used the synthetic range data with added noise and second we used real range data scanned with a commercially available range sensor from SICKIVP. The synthetic 'blox' image, seen in figure 1, has 400^2 measure points and was voxelized into a volume of 256^3 voxels. In the synthetic

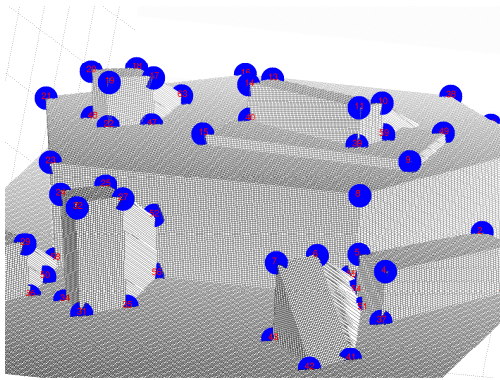


Figure 1. Zoom in on synthetic version of 'blox' with handmarked corners.

data we handmarked 67 corners and then ran the detectors, equations (7)-(13), and measured correctly detected corner and false positives. These tests were performed for noise-free images as well as for 5 levels of added noise. The results can be seen in tables 2-3.

dB:	4	12	20	31	44	Inf
D1	54.2	51.6	52	51.4	52.4	54
D2	53.4	50.6	51.4	51	51	55
D3	51.8	50	50	48.8	49.4	51
D4	51.8	49.8	50.2	48.8	49.4	52
Op3	55.4	54	55.8	55.6	56.2	57
Rohr3D	54.8	52.6	55.4	54.8	56	56
Laptev	53.8	51.4	51.8	51.2	52.2	56

Table 2. Correctly found corners per SNR.

From tables 2-3 we see that the all the detectors find roughly the same amount of corners. We also see that detectors which use only the third degree invariant are

dB:	4	12	20	31	44	Inf
D1	166.6	68.2	25.8	13.2	12.4	16
D2	151.6	65	23.4	12.4	11	13
D3	126.2	57	17.6	11	10.6	12
D4	128.8	57.8	18.4	11	10.6	12
Op3	199.8	103.6	53.6	27.4	25	25
Rohr3D	188.8	97	48.8	25	23.4	27
Laptev	146.6	72.8	29	13.8	14.2	14

Table 3. False positives per SNR.

the most prone to find false positives, i.e. corners we did not categorize as corners. Detectors using second and first degree invariants are less likely to find false positives and less prone to do that as noise increases. In this test, the detectors using all three invariants are the most stable ones.

For the real range data we ran the best detectors and visually inspected the results for some different scanning directions of a few objects. The implemented detec-

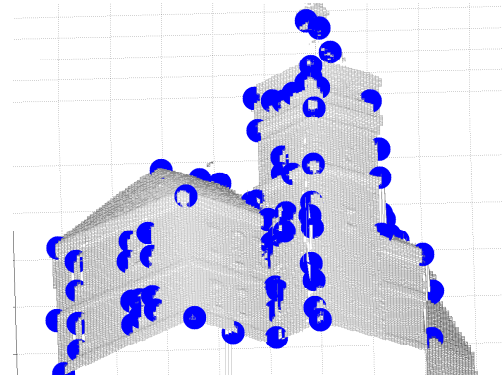


Figure 2. A scan of a miniature house with POIs detected by D3.

tors do detect POIs on real range data, see 2, but more testing is needed. From the visual inspection of POIs detected by the different detector variants it seems the results on real data are analogous with the results on synthetic data; using all three invariants of the Harris matrix results in a more stable POI-detection.

5. Conclusions and Future Work

We have seen that differential-based corner detectors, including previously published ones originally developed for full-volume data, such as MRI data, do work for range data if a conservative surface voxelization step is performed. From our comparison we see that using all three invariants of the Harris matrix in the

detector can be beneficial for range data applications. Eventhough results of running differential-based detectors using voxelized data do look promising, there is still some work to be done. We would like to extend our voxelization to a sub-voxel scan-conversion to see how this might improve the performance of the detector. We do need to add sub-voxel refinement of the localization of the POI and we would like to add a scale-space to our detector implementation. When that work finishes a new evaluation on real range data were POIs are tracked under rotation would be fitting. As a final thought one has to keep in mind that a Harris-based detector are really only suitable for range data of objects where a high degree of non-smoother surface-parts do exist.

6. Acknowledgements

This research has in part been funded by SICKIVP.

References

- [1] J. R. Alzola, R. Kikinis, and C. Westin. Detection of point landmarks in multidimensional tensor data. *Signal Processing*, 81:2243–2247, 2001.
- [2] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded-up robust features. In *ECCV*, 2006.
- [3] H. Chen and B. Bhanu. 3d free-form object recognition in range images using local surface patches. In *ICPR (3)*, pages 136–139, 2004.
- [4] C. S. Chua and R. Jarvis. Point signatures: A new representation for 3d object recognition. *Int. J. Comput. Vision*, 25(1):63–85, 1997.
- [5] R. S. Correa, L. G. Shapiro, and M. Melia. A new signature-based method for efficient 3-d object recognition. *CVPR: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1, 2001.
- [6] Z. Dong, W. Chen, H. Bao, H. Zhang, and Q. Peng. Real-time voxelization for complex polygonal models. In *PG '04: Proceedings of the Computer Graphics and Applications, 12th Pacific Conference*, pages 43–50, Washington, DC, USA, 2004. IEEE Computer Society.
- [7] A. Flint, A. Dick, and A. van den Hengel. Thrift: Local 3d structure recognition. In *DICTA '07: Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, pages 182–188, Washington, DC, USA, 2007. IEEE Computer Society.
- [8] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Malik. Recognizing objects in range data using regional point descriptors, 2004.
- [9] C. G. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference*, pages 147–151, September 1988.
- [10] T. Hartkens, K. Rohr, and H. Stiehl. Evaluation of 3D Operators for the Detection of Anatomical Point Landmarks in MR and CT Images. *Computer Vision and Image Understanding*, 86(2):118–136, 2002.
- [11] G. Hetzel, B. Leibe, P. Levi, and B. Schiele. 3d object recognition from range images using local feature histograms.
- [12] A. E. Johnson and M. Hebert. Recognizing objects by matching oriented points. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 684, Washington, DC, USA, 1997. IEEE Computer Society.
- [13] I. Laptev. On space-time interest points. *Int. J. Computer Vision*, 64(2):107–123, 2005.
- [14] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.
- [15] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, volume 1, pages 384–393, 2002.
- [16] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. 60(1):63–86, 2004.
- [17] H. P. Moravec. Visual mapping by a robot rover. In *Int. Joint Conf. on Artificial Intelligence*, pages 599–601, August 1979.
- [18] R. Pielot, E. D. Gundelfinger, M. Scholz, K. Obermayer, and A. Hess. Performance of 3d landmark detection methods for point-based warping in autoradiographic brain imaging. In *SSIAI '02: Proceedings of the Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, page 269, Washington, DC, USA, 2002. IEEE Computer Society.
- [19] K. Rohr. On 3D Differential Operators for Detecting Point Landmarks. *Image and Vision Computing*, 15(3):219–233, 1997.
- [20] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. 19(5):530–535, 1997.
- [21] F. Stein and G. Medioni. Structural indexing: Efficient 2d object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(12):1198–1204, 1992.
- [22] Y. Sun and M. A. Abidi. Surface matching by 3d point's fingerprint. In *ICCV*, pages 263–269, 2001.
- [23] J.-P. Thirion. Extremal points : definition and application to 3D image registration. In *IEEE conf. on Computer Vision and Pattern Recognition*, Seattle, June 1994.
- [24] F. Viksten and K. Nordberg. A geometry-based local descriptor for range data. In *DICTA '07: Proceedings of the 9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, pages 210–217, Washington, DC, USA, 2007. IEEE Computer Society.
- [25] S. M. Yamany and A. A. Farag. Free-form surface registration using surface signatures. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1098, Washington, DC, USA, 1999. IEEE Computer Society.
- [26] L. Zhang, W. Chen, D. S. Ebert, and Q. Peng. Conservative voxelization. *Vis. Comput.*, 23(9):783–792, 2007.
- [27] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI*, 78(1-2):87–119, 1995.