Regular Polygon Detection as an Interest Point Operator for SLAM

David Shaw[†‡], Nick Barnes[‡]

† Department of Systems Engineering
Research School of Information Sciences and Engineering
Australian National University

‡ National ICT Australia
Locked Bag 8001 Canberra ACT 2601

davids@syseng.anu.edu.au, nick.barnes@nicta.com.au

Abstract

We present a new interest point operator based on the regular polygon detector developed by Loy and Barnes [2004]. This operator finds square-like features as a basis for scene reconstruction and visual Simultaneous Localisation and Mapping (SLAM) from robot camera sequences. In this paper we show results from the application of this detector as an interest point operator on a robot camera sequence in an indoor office environment. The detector shows good results for non-trivial frame baselines.

1 Introduction

Reconstructing the elements within a video sequence is an aspect of computer vision that has application in visual SLAM (Simultaneous Localisation And Mapping) for robotics. An important aspect of scene reconstruction and image analysis is the interest point operator; a good interest point operator should detect the same feature present in a scene from different camera positions with high repeatability, This results in high quality point correlation matches between two video frames. Typical interest point operators, such as the Harris corner detector [Harris and Stephens, 1988], detect point representations of corners.

In their work on road sign detection, Loy and Barnes outlined their regular polygon detector, which detects equiangular polygons based on their inherent properties of symmetry[Loy and Barnes, 2004]. The radial symmetry operator for circle detection[Loy and Zelinsky, 2003; Barnes and Zelinsky, 2004] was extended to detect regular polygons by extending the vote from each pixel along a region of possible centres perpendicular to the gradient of the pixel. This algorithm has been shown to be effective on detecting polygonal road signs, such as the regular triangular, square and octagonal shapes prevalent in use. In the case of detecting landmarks such as road signs, the shapes are well distinguished from the background, and usually orthogonal to the direction of traffic. This detector can provide the position of the polygon as well as the radius and shape rotation angle.

In this paper, we introduce a new interest point operator based on the regular polygon detector that is presently under development for use in feature extraction for scene reconstruction and robot SLAM. As shown by Loy and Barnes [2004], the regular polygon detector is well suited for road sign detection, however the presence of such well defined regular polygons in a more general environment is not as guaranteed. This paper outlines the use of this polygon shape detector in order to find square based interest points, for application in robotic vision as part of localisation and scene reconstruction. The use of the shape detector for the locating of road signs required a reasonable threshold for the sum of votes from pixels to signal the presence of a polygon. By lowering these thresholds the detector can be used to locate features within an image that have part of the qualities of a polygon, such as corners or parallel lines, that can be reliably located in similar images, such as those in close proximity frames from on-board robot camera video. Finding square-based features has the advantage of providing many dimensions of information per shape, such as the radius, rotation angle and a measure of 'squareness'. Square-based features also possess the property of being robust to changes in scale in the image.

2 Algorithm

The regular polygon detection algorithm is described in detail by Loy and Barnes [2004], however for the sake of completeness in this paper the algorithm will be outlined here.

The regular polygon detector is an extention of the fast radial symmetry transform[Loy and Zelinsky, 2003], which detects circles of specified radii. It operates on the gradient of the gray-scale of an image. Let G denote the gradient elements of the image, i.e. G is a vector field with two components G_x, G_y.

Only the gradient elements above a threshold (designated β) are considered by the algorithm; these are normalised, with the other elements set to zero. i.e. let $G_\beta = \{\hat{\mathbf{g}} = \frac{\mathbf{g}}{\|\mathbf{g}\|}, \mathbf{g} \in G \mid \|\mathbf{g}\| > \beta\}$

Each of these remaining normalised gradient elements vote for the potential circle centre for a given radius r along the direction of the normalised vector $\hat{\mathbf{g}}$. The votes are tabulated into another vector field, where the total score for potential circle centre pixel $\mathbf{c}$ is the sum of all vectors $\hat{\mathbf{g}} \in G_\beta$ for which $\mathbf{c}$ is the circle centre point, or *affected pixel*, and is defined by:

$\mathbf{c}(\mathbf{p_c}) = \mathbf{c}(\mathbf{p}) + \mathbf{g}(\mathbf{p})$, where $\mathbf{p_c} = \mathbf{p} + r\mathbf{g}$.

Since it is possible to have either light shapes on a dark background or vice versa, two circle centres are voted for.

$\mathbf{c}_{\pm ve}(\mathbf{p}) = \mathbf{p} \pm \text{round}\ (r\mathbf{g}(\mathbf{p}))$

For circle detection, the affected pixel vote score can be calculated by tallying up the votes for it, the score is just the number of votes cast. The affected pixels with the highest score are the most likely candidates for circles of the specified radius.

This algorithm is extended to detect regular polygons. Here the radius is defined as the perpendicular distance from the edge to the centroid of the shape ($\mathbf{c}$ of the polygon). Since the direction to the centroid is not always directly perpendicular to the tangent of the perimeter of the shape, the voting system must be extended to account for this. A line of votes is required for all the possible centroid positions, cast for affected pixels on a line segment (parallel to the edge) perpendicular to the vector. The length of the vote line segment differs for each type of regular polygon to be detected, given by the formula $w = \text{round}\ (r \tan \frac{\pi}{n})$; for squares, $w = r$.

To detect regular polygons with the appropriate number of sides the nature of voting for affected pixels is changed to accommodate this. Since the angles of regular polygons are all congruent, this can be used to form a rotationally invariant method of detecting this type of polygon. The method is to perform a transformation on the vector to be added to the affected pixels.

If $\mathbf{v} = (1, \theta)$, where (r, θ) is the polar coordinates of the vector (note that $r = 1$ as $\mathbf{v}$ is normalised), then let $\mathbf{u} = (1, n\theta)$, where n is the number of sides of the regular polygon being detected. For a regular polygon with n sides, the possible gradient vectors are of the form $\mathbf{v} = (1, \alpha + \frac{2\pi k}{n})$, where $k \in [0, n-1]$. With this transform this becomes $\mathbf{u} = (1, \alpha + n\frac{2\pi k}{n}) = (1, \alpha + 2\pi k) = (1, \alpha)$ for all the points in the perimeter of the polygon. The sum of these vectors will be maximal at the centroid of the polygon. Good candidates for polygon centroids will have many such adjusted gradient vectors summed at their pixel position in the vote vector field, and as such the result will have a large magnitude.

This vector sum can also be used to detect the angle of rotation, as for strong candidates a large number of voting pixels from the perimeter of the polygon will have the appropriately angled vector. With a large proportion of perimeter pixels to noise the resulting vector vote will give a good measure on the polygon angle.

1. Determine the gradient vector field. Threshold the magnitude, setting values below the threshold to zero and those above to unity. Denote the output $\mathbf{g}$.
2. Determine the n-angle gradient such that $\|\mathbf{v}\| = \|\mathbf{g}\|$ and $\angle\mathbf{v} = n\angle\mathbf{g}$.
3. For each radius under consideration:
   (a) Consider each non-zero element of $\mathbf{g}$ in turn, for each such element:
      i. Determine the vote locations.
      ii. Accumulate the contribution to the vote image O_r, and the equiangular image B_r.
   (b) Calculate the output image S_r at radius r, as $O_r(\mathbf{p})B_r(\mathbf{p})$, and accommodate for scale.
4. Sum S_r over all radii $r \in \mathcal{R}$ to determine the final output image S.

Figure 1: Summary of the regular polygon detector. Reproduced from paper by Loy and Barnes [2004].

3 Implementation

In order to be used in generalised indoor environments, some adaptions were required to the regular polygon detector for use as a interest point operator. Since the typical indoor environment does not include many regular polygons, the detector is used to detect polygonal features. In this case, due to the rectilinear nature of man-made environments, the regular polygon detector is used to detect square-like features, such as corners. As these features are common in indoor environments, and are re-detectable with small baseline changes in camera position, they are suitable candidates for detection. This detection process works as the voting system will give reasonable scores to features that only have some of the structure of a square, such as only two sides of the shape. Thus persistent features such as right-angled corners or parallel lines will get half the score as a properly realised square.

We tested the regular polygon feature detector on a robot camera sequence. The sequence consists of the camera input of a robot travelling in a closed loop in corridors around a typical indoor office environment. The thresholds for these tests have been empirically set: fine-tuning of these thresholds is still an ongoing considera-

tion. For this test the cut-off threshold for the gradient vector field (β) is set to 0.125, and the shape detection threshold for the voting system (σ) is set to 0.667. The detector is set to look for squares, with a range of only three radii: $4.5, 6.75$ and 10.125.

The regular polygon feature detector is run on frames from the video sequence to find a set of potential squares within the frame. The detector provides the centroid coordinates, radius, rotation angle and strength of the vote (a measure of the 'squareness', i.e. the degree of how much the feature is like a square). The detected squares are then matched using a correlation comparison of the local neighbourhood of the centroid of each square. An advantage presented in this step by using the regular polygon detector over more conventional point corner detectors is the extra information provided can be used as an aid, assuming a reasonably small baseline between frames. In this experiment only pairs of squares with a rotation angle difference of less than $\frac{\pi}{16}$ radians are compared when correspondences are found in the correlation stage.

4 Results

In order to test the effectiveness of this interest point operator for a sample pair of frames on this sequence, the pair correspondences between squares in two frames are compared with a fundamental matrix found independently. The measure used (adapted from similar approaches used for other interest point detectors[Mikolajczyk and Schmid, 2003]) is to consider a pair correspondence a match if the closest geometric distance between a point and the expected epipolar line is less than a specified threshold (5 pixels in this test).

For longer sequences of comparisons it is not practical to find independent fundamental matrices. Thus a further test used is to calculate a fundamental matrix using a limited form of RANSAC and comparing the output with the matches found. A high number of match points suitable with the found fundamental matrix is a fair indicator of a good data set, while a low number is a certain indicator of a bad data set.

4.1 Image Pair Analysis

As an example of the application of the detector, a sample pair of frames are taken from the robot camera sequence, shown in figures 2(a) and 2(b). These two images are separated by 10 frames, or around 3.3 seconds in time, so that although reasonably close in the sequence, it can be seen that the robot has moved a significant distance between the two. The gradient vector field calculated for these two frames is shown in figures 3(a) and 3(b).

After performing the regular polygon detector on these frames, the squares found are shown in Figures 4(a) and 4(b) respectively. Based on the strength of the vote the squares are coloured from red to yellow (dark grey to light grey), from low strength to high. Each square found has five dimensions of information: the 2 dimensional coordinates of the centroid of the square, the radius, the rotation angle and the strength of the vote (representing the squareness of the feature).

As can been seen in these images, many of the strong square candidates arise from parallel lines of the appropriate radius. As a feature found with the regular polygon detector they have the disadvantage of being variable; without a texture to influence the vote the spacing of the squares along these parallel lines are somewhat arbitrary. The counterbalance to this is the parallel lines themselves are very persistent to detection, and the large number of squares found will ensure that some pairs of squares will be matched in the correlation stage.

The found square features are matched between the two frames by correlation, comparing the local neighbourhood of image around the centroid of the shapes. This step is improved by using the extra information present in the shape data. As only three radii are being searched for in these examples, the squares are only matched between those of the same radius. The matches are also decided by the similarity of their rotation and vote strength. Figure 5 shows the pair correspondences that were considered matches between squares in these frames. The dark lines indicate good matches, the light lines indicate erroneous matches. Good matches are decided by comparing the position of the points with the expected epipolar line from a fundamental matrix determined from an independent set of reliable data. This measure has some degree of error as erroneous points may fall close to the epipolar line, however it acts as a reasonable indicator of the correctness of the data set. In this particular case, the pillars and door at the end of the corridor have a high match correspondence, together with the light fixtures in the ceiling. There are a high proportion of erroneous matches with the doors on the edge of the frames due to the particular aliasing effect as the baseline chosen with the speed of the robot happens to correlate to the spacing of the doors. Nevertheless, 43% of the pairs were considered good by the metric given above.

As a further test of the data, a fundamental matrix was generated with a RANSAC algorithm (using the techniques given by Hartley and Zisserman [2000]), limited to 1000 random data sets. Figure 6 shows some sample epipolar lines with the epipole at their intersection generated by this method.

(a) First frame

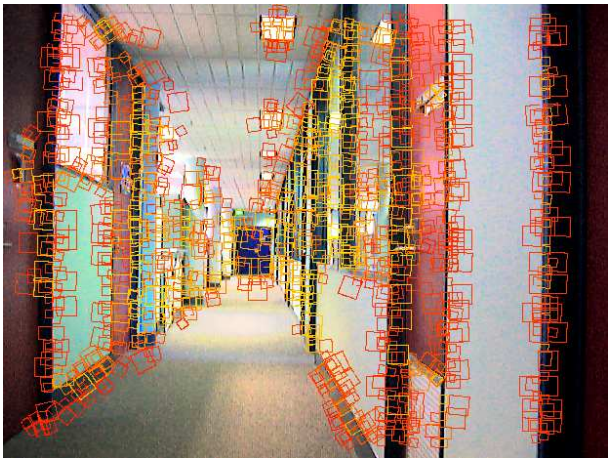
(b) Second frame

Figure 2: Frames for comparison

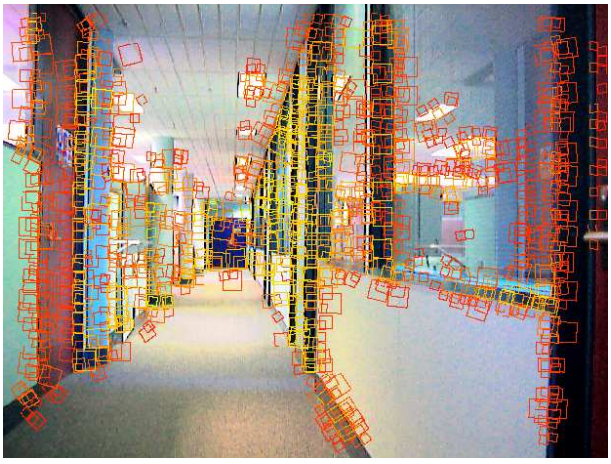
(a) First frame

(b) Second frame

Figure 3: Gradient image from frames



(a) First frame



(b) Second frame

Figure 4: Squares detected in frames

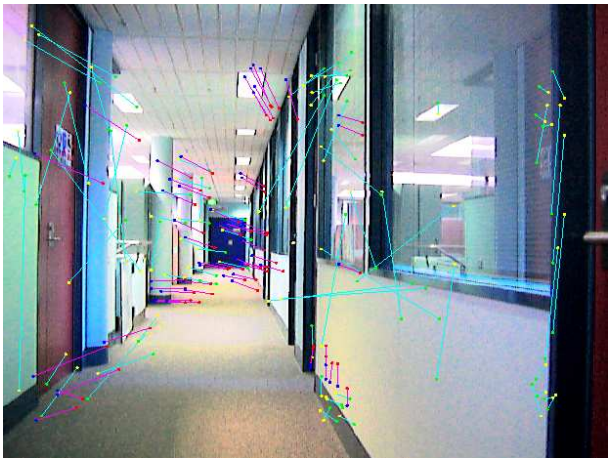


Figure 5: Pair correspondence matches between squares in frames

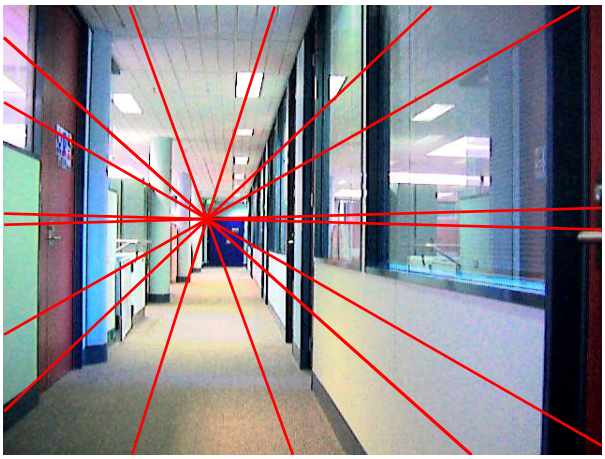


Figure 6: Projected epipolar lines

4.2 Sequence Analysis

The frame matching process described above is applied to the entire robot camera sequence. The step of 10 frames (3.3 seconds) is used as a non-trivial baseline between images. The resulting data is summarised in figures 8, 9 and 10.

Figure 8 shows the number of squares found in each frame. Over 700 squares were found in all frames with one exception (a relatively featureless frame captured when the robot is turning, see figure 7.

Figure 9 shows the number of pair correspondences found between sequential candidate frames that were considered to be matches by the algorithm. There are a reasonable number of matches found for the periods when the robot is travelling straight down the corridor. The periods with low matches correspond to the frames when the robot is turning which is unavoidable, as the robot can turn through 90° in 10 frames. The frames with high numbers of matches arise from when the robot has stopped or hardly moving.

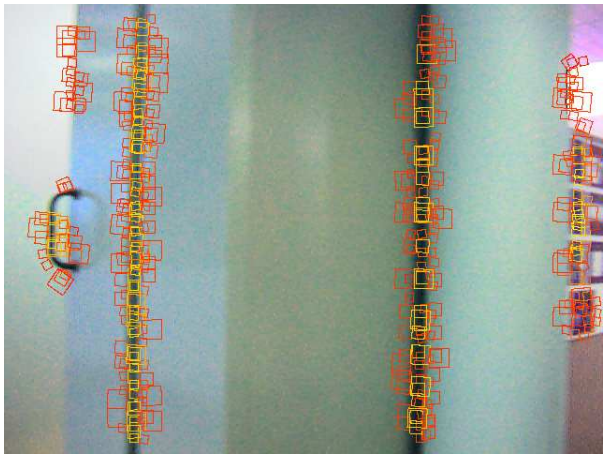


Figure 7: Sample bad frame to detect squares

Figure 10 shows the fraction of pair matches that were actually considered matches in the calculation of the final fundamental matrix decided by RANSAC. As can be seen comparing this with figure 9, the fraction of good matches is correlated with the number of matches found.

Figure 11 shows six sequential candidate frames during a straight corridor section of the camera sequence, with the epipole marked by a cross. While some variance is present from the true epipole of motion, it shows reasonable accuracy for the degree of variance between the frames.

4.3 Present Limitations

One limitation with this detector that is still present at this time is the necessary dependence on well-defined edges for the shape detector. As can be seen in figures 4(a), 4(b) and in particular figure 7, squares will not be found in areas without a significantly sharp intensity change. While featureless blank areas cannot be expected to provide much information for an interest point operator, more faintly defined lines such as the tiling in the ceiling could be used as a source for potential squares. This problem can be addressed by adjusting the gradient threshold β used in the operation of the detector; at present determining the optimal values of these thresholds is yet to be completed.

5 Conclusion

We have presented our interest point operator based on regular polygon detection, which has been demonstrated over a robot camera sequence. This method has been shown to give good performance on real robot camera data with a non-trivial baseline between camera frames.

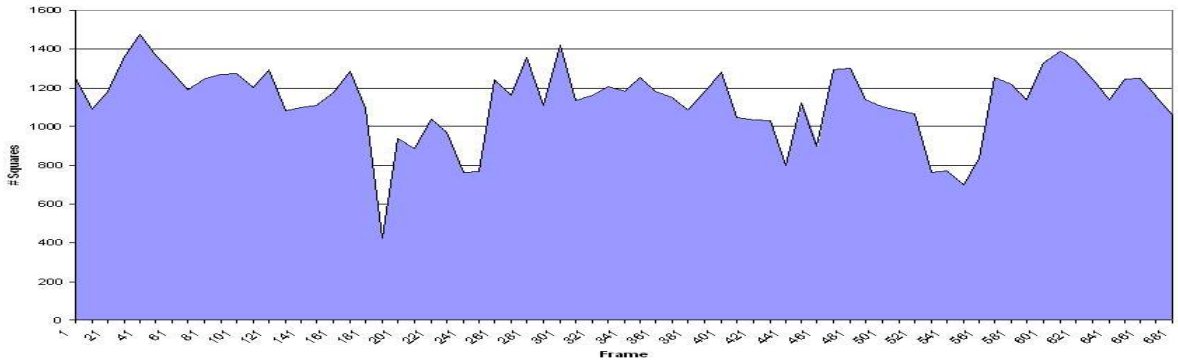


Figure 8: Number of squares detected

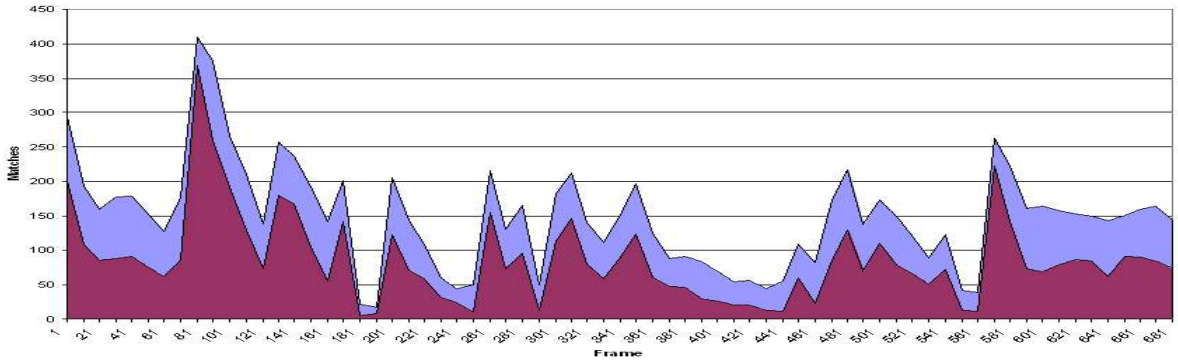


Figure 9: Light region (blue): Number of pair correspondence matches between squares
Dark region (red): Number of good matches as determined by fundamental matrix

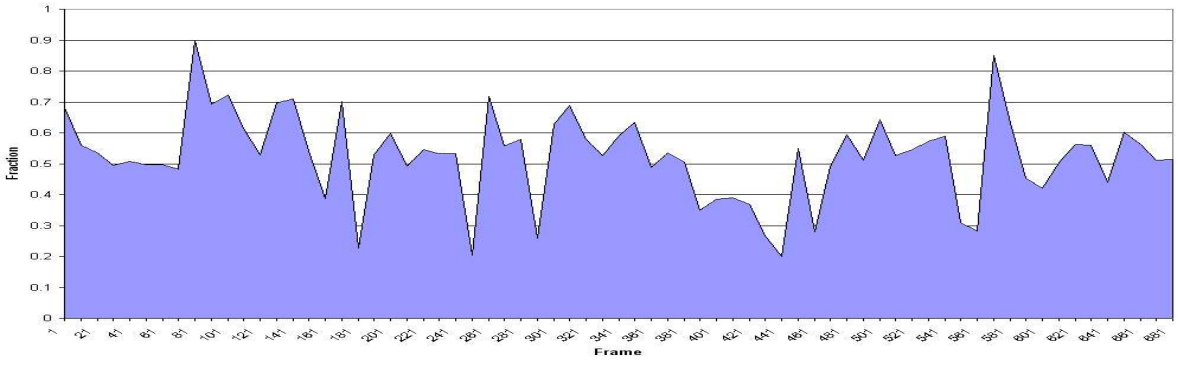


Figure 10: Fraction of matches fitting expected epipolar geometry from calculated fundamental matrix

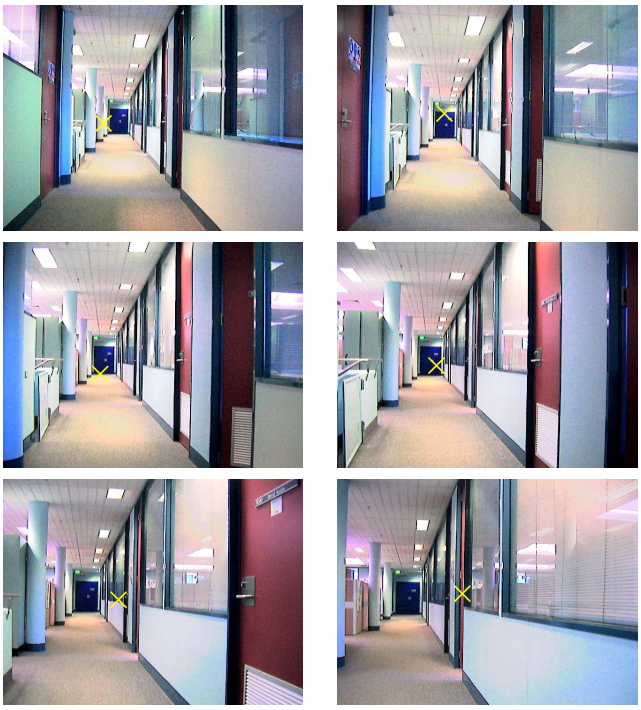


Figure 11: Frames from robot camera sequence showing epipole

We feel that the regular polygon detector shows strong potential to be used as the basis of an general interest point operator for use in robot localisation and mapping.

References

N. Barnes and A. Zelinsky. Real-time radial symmetry for speed sign detection. In *Proceedings IEEE Intelligence Vehicles Symposium*, 2004.

C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

R. Hartley and A. Zisserman. *Multiple View Geometry in computer vision*. Cambridge University Press, 2000.

G. Loy and N. Barnes. Fast shape-based road sign detection for a driver assistance system. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, 2004.

G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 25, pages 959–979, August 2003.

K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition*, June 2003.