Using Local Symmetry for Landmark Selection

Gert Kootstra[1], Sjoerd de Jong, and Lambert R.B. Schomaker

University of Groningen, Postbus 407, 9700 AK Groningen, The Netherlands,
G.Kootstra@ai.rug.nl, WWW home page: http://www.ai.rug.nl/ gert

Abstract. Most visual Simultaneous Localization And Mapping (SLAM) methods use interest points as landmarks in their maps of the environment. Often the interest points are detected using contrast features, for instance those of the Scale Invariant Feature Transform (SIFT). The SIFT interest points, however, have problems with stability, and noise robustness. Taking our inspiration from human vision, we therefore propose the use of local symmetry to select interest points. Our method, the MUlti-scale Symmetry Transform (MUST), was tested on a robot-generated database including ground-truth information to quantify SLAM performance. We show that interest points selected using symmetry are more robust to noise and contrast manipulations, have a slightly better repeatability, and above all, result in better overall SLAM performance.

1 Introduction

One of the fundamental tasks of an autonomous robot is to build a map of the environment and use it for self localization. The problem of Simultaneous Localization and Mapping (SLAM) has therefore received much attention in the last decade [1]. Nowadays, approaches using laser range finders are very successful. SLAM using vision, however, remains a challenging research topic, e.g., [2, 3].

Using a camera has the advantage over a laser-range finder that it is a passive sensor that is low cost, low power, and lightweight. A camera furthermore provides a rich source of information, which enables the use of sophisticated detection and recognition methods. The difficulty, however, is to extract relevant information from the high-dimensional visual data in real time. In this paper, we propose the use of local symmetry.

Most visual SLAM systems use visual landmarks to create a map of the robot's environment. It is important to select robust, stable landmarks that will be recognizable in future encounters. Furthermore, the number of selected landmarks should be limited, since the computational complexity of the SLAM algorithms strongly depends on the number of landmarks.

Usually, interest-point detectors are used for landmark selection. Most approaches use contrast features to detect interest points. Examples are the Scale-Invariant Feature Transform (SIFT) [4, 5], Speeded Up Robust Features (SURF) [6, 7], and Harris corners [8]. We propose the use of local symmetry to detect interest points. We compare our method with SIFT, since it is among the best performing interest point detectors in SLAM [9], as well as in object recognition [10].

Although systems using SIFT have successfully been applied to SLAM, there are three important drawbacks. The interest points are very susceptible to noise, not all selected landmarks are recognized when the robot returns to a previously visited location,
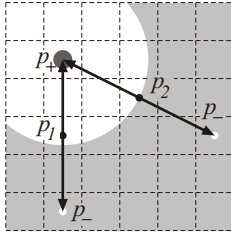
Fig. 1. The Fast Radial Symmetry Transform [21]. Each pixel in the image votes for symmetry at a given radius r. In this example, two pixels p_1 and p_2 are shown. Using the orientation of a pixel's gradient, a vote is made for bright symmetrical forms on a dark background at p_-, and the reverse, p_+. Doing this for all pixels in the image gives the symmetry transform of the image, $\psi_r(o, s)$, at octave o and scale s.

and too many interest points are found in an image, but only few points are so stable that they can be tracked over a number of successive frames. Thus SIFT has problems with robustness, repeatability, and stability.

We propose a landmark selection mechanism using local symmetries instead of local contrast, and show that it performs better regarding the above mentioned problems. Our choice for symmetry is motivated by human behavior. Symmetry is detected very rapidly, especially when patterns have multiple axes of symmetry [11]. Moreover, humans pay attention to locally symmetric parts of images [12, 13]. Furthermore, humans use symmetry to segregate a figure from its background [14]. These findings suggest that symmetry is used in pre-attentive vision, and can therefore be used for context-free object segmentation and landmark selection.

Using symmetry to select landmarks exploits the fact that most man-made indoor environments contain many symmetrical objects and forms. Since symmetry is a strong non-accidental property, we believe that its use will result in the selection of valuable landmarks for the visual SLAM system.

Although contrast features have received most attention in computer vision research (e.g., [4, 15]), symmetry is successfully used in a number of studies. In earlier work, for instance, Marola [16] used symmetry for detection and localization of objects in planar images. Symmetry has furthermore been used to control the gaze of artificial vision systems [17, 18] Heidemann [19] showed that interest points detected with color symmetry are robust to noise and 3D object rotation. Moreover, he showed that symmetry detection results in points that are more robust to changing light conditions than Harris corners and that these points are more unique compared to other locations in the image.

To detect local symmetry in an image, a number of symmetry operators exist. Reisfeld *et al.* [20], for instance, developed a mirror-symmetry operator by comparing the gradients of neighboring pixels. Heidemann [19] extended this work to the color domain. Reisfeld *et al.* also proposed a radial-symmetry operator that promotes patterns that are symmetric in multiple symmetry axes. A faster operator to detect radial symmetry, the Fast Radial Symmetry Transform (FRST), is proposed in [21].

In this paper, we use FRST as a basis of our model to determine local symmetry. We develop the model to a novel scale- and rotation-invariant interest-point detector, which we call the MUlti-scale Symmetry Transform (MUST). We discuss the use of MUST for selecting landmarks for visual SLAM.

We show that landmarks selected using local symmetry are more robust to noise and have a higher repeatability, and require fewer computations. Most importantly, we

show that the overall performance of the SLAM system increases when interest points are selected using local symmetry instead of SIFT.

2 Methods

In this section we first explain the selection of interest points using local symmetry. We then discuss the visual buffer to select stable interest points as landmarks. We end with a description of the Kalman Filter implementation for visual SLAM.

2.1 Interest Points Based on Local Symmetry

As a basis of our interest point detector, we use the Fast Radial Symmetry Transform [21], and extended it to a multi-scale and rotational-invariant detector, MUST. The basis of the FRST is given in Fig. 1, and MUST is depicted in Fig. 2.

To obtain a multi-scale interest-point detector, we use a pyramid approach similar to that used in [4]. The symmetry response is calculated at five spatial octaves, $\mathcal{O} = \{-1, 0, 1, 2, 3\}$. In the first octave, -1, the gray-scaled input image, I_{-1}, has twice the resolution of the original image, similar to [4]. For each next octave, the input image of the previous octave is smoothed with a Gaussian kernel and down-sampled by a factor of two. Within each octave, there are three scales, $s \in \{0, 1, 2\}$, with progressive smoothing. This gives a pyramid of gray-scaled images, $I_{o,s}$, for a given octave o and scale s.

The symmetry transform, $\Psi(o, s)$, for octave o and scale s, is calculated by investigating the local radial symmetry at a set of different radii. The size of the radii depends on the scale s. The set of radii used to calculate the symmetry response is defined as $\mathcal{R}_s = (1 + s/(s_{max} - 1)) \cdot \mathcal{R}$, where $s_{max} = 3$ and $\mathcal{R} = \{1, 3, 5\}$. The symmetry transform is then:

$$\Psi(o, s) = \frac{1}{|\mathcal{R}_s|} \sum_{r \in \mathcal{R}_s} \psi_r(o, s) \tag{1}$$

where $\psi_r(o, s)$ is the symmetry response at octave o and scale s for radius r.

$\psi_r(o, s)$ is determined by first calculating the gradients of the input image $I_{o,s}$ with horizontally and vertically aligned Sobel filters, resulting in a magnitude map, $G_{o,s}$, and an orientation map, $\Theta_{o,s}$. Then, each pixel p votes for the existence of symmetry based on its gradient, $\Theta_{o,s}(p)$, its magnitude, $G_{o,s}(p)$ and the given radius r. This is related to the Hough transform. A pixel votes for the existence of both a bright symmetrical form on a dark background, and a dark symmetrical form on a bright background at respectively location p_+ and p_-:
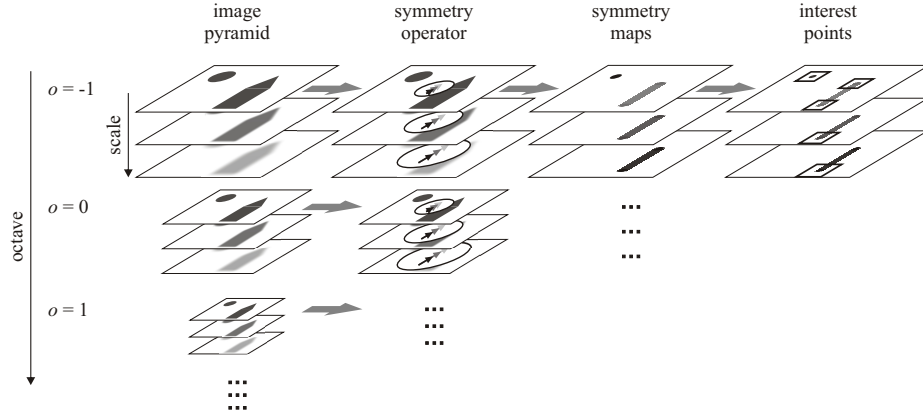
$$p_+ = p + r \cdot (\cos \Theta_{o,s}(p), \sin \Theta_{o,s}(p)) \tag{2}$$
$$p_- = p - r \cdot (\cos \Theta_{o,s}(p), \sin \Theta_{o,s}(p)) \tag{3}$$

where p_+ and p_- are rounded to the nearest pixel. Subsequently, an orientation projection map, O_r, and a magnitude projection map, M_r, are built. Initially, all values in these maps are set to zero. Then, for every pixel p, the maps are updated according to:

$$O_r(p_+) = O_r(p_+) + 1 \, , \, O_r(p_-) = O_r(p_-) - 1 \tag{4}$$
$$M_r(p_+) = M_r(p_+) + G_{o,s}(p) \, , \, M_r(p_-) = M_r(p_-) - G_{o,s}(p) \tag{5}$$

Fig. 2. MUST, the MUlti-scale Symmetry Transform. Symmetry is calculated at multiple octaves and scales. At the first octave, $o = -1$, the image is double its original size. For every next octave, the image is down-scaled by a factor two. Within an octave there are three scales with progressive Gaussian blurring. For every image in the image pyramid, the symmetry response, $\psi_r(o, s)$, is calculated at three different radii by applying the symmetry operator. The average of these responses results in the symmetry map, $\Psi(o, s)$. In the last step, the interest points are obtained by finding points that have a local maximal or minimal symmetry value. The local neighborhood of such an interest point is described by the SIFT descriptor.

Finally, the symmetry map for radius r at octave o and scale s is determined by:

$$\psi_r(o, s) = F_r * A_r \tag{6}$$

and

$$F_r(p) = M_r(p) \cdot O_r(p)/k_r \tag{7}$$

where $O_r(p)$ has a upper value of k_r, and a lower value of $-k_r$. k_r has been experimentally established in [21] at $k_r = 8$ for $r = 1$, and $k_r = 9.9$ otherwise. A_r is a Gaussian kernel of size $r \times r$ with a standard deviation of $0.25r$.

This gives us the symmetry response $\psi_r(o, s)$ for the radius r at octave o and scale s. By averaging the symmetry responses over all radii in $\mathcal{R}_s$, according to equation (1), we obtain the full symmetry response at octave o and scale s, $\Psi(o, s)$.

Next, the interest points in every octave and scale are determined by finding the points in $\Psi(o, s)$ that are either a maximum or a minimum in the neighborhood of 11×11 pixels. Each interest point i found in the symmetry maps is stored. We store the following information: (a) the location of the interest point in the original resolution of the input image, $\mathbf{x}_i = (x_i, y_i)$, (b) its scale value, $\sigma_i = 2^{o_i + s_i/s_{max}}$, (c) its symmetry strength, $v_i = \Psi(o_i, s_i, \mathbf{x}_i)$, (d) its orientation, γ_i, and (e) the descriptor of the interest point, $\mathbf{d}_i$. The later two are described in the next paragraph.

To calculate the orientation and the descriptor of the interest points, we use the corresponding methods of the SIFT algorithm [4]. The orientation of the interest point is determined by finding the dominant gradient in the local neighborhood of the interest
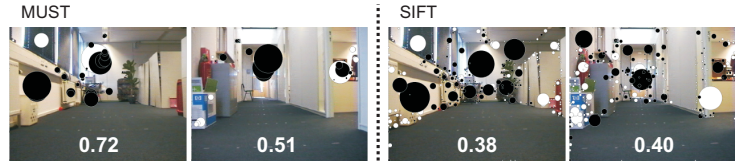
Fig. 3. Two examples images with the MUST and SIFT interest points. The black circles present the stable interest points that are found in both the current and the previous image. The white circles depict the unstable interest points that are only found in the current image. The proportion of stable points is given at the bottom.

point. This orientation is used to obtain a rotationally invariant descriptor of the local neighborhood patch. The neighborhood is described by histograms of gradients. The size of the neighborhood depends on the scale value of the interest point. This makes that the descriptor is also scale invariant. To calculate the histograms of gradients. the patch is divided in 4 by 4 squares. A histogram of gradients is then calculated for each square. Since there are 16 squares, and each histogram contains 8 bins, this gives us a feature vector with 128 values. The magnitude of the feature vector is normalized to 512. For more detailed information about the method to calculate the descriptor, we refer to [4].

Figure 3 shows the interest points found by MUST on two of the images used in our experiments as compared to the SIFT interest points. Note that SIFT results in a large number of interest points, of which many are unstable. MUST, on the other hand, results in less, but more stable points.

2.2 The Visual Buffer

Both MUST and SIFT result in a large number of interest points when applied to the camera images taken by the robot. In our experiments, MUST detects on average 40 interest points per image, and SIFT 124. To be practically usable, the state matrix of the Extended Kalman Filter should contain a few hundred landmarks in total. Furthermore, most interest points are unstable, and therefore useless for SLAM. We propose a visual buffer test the stability of the interest points, similar to [2, 5].

The buffer contains the last N camera images. The interest points in the current image are compared to those in the $N-1$ previous images. An interest point i passes the buffer if it meets two criteria: 1) The point matches in at least M of the previous images. Two points, i and j, match when the descriptors, $\mathbf{d}_i$ and $\mathbf{d}_j$, are sufficiently similar. This is true when the Euclidean distance is below the threshold τ_1: $\|\mathbf{d}_i - \mathbf{d}_j\| < \tau_1$, and when the *best-to-next-best ratio* is smaller than the threshold δ_1: $\|\mathbf{d}_i - \mathbf{d}_j\|/\|\mathbf{d}_i - \mathbf{d}_l\| < \delta_1$, where $\mathbf{d}_l$ is the descriptor of the second most-similar region in the previous image. This ratio ensures uniqueness. Additionally, to avoid spurious interest points, the symmetry strength of the interest point should be $v_i \geq \lambda \cdot v_{max}$, where v_{max} is the strength of the most symmetrical point that is successfully matched in the current image, 2) The estimates of the position of the landmark in the environment are congruent. The landmark's position is estimated by triangulation using the bearings of the interest point

in the current image and the previous images, and the displacement of the robot. This results in a set of estimates of the range and bearing, $\mathrm{P} = \{\mathbf{p}_k | \mathbf{p}_k = \langle r_k, \theta_k \rangle \wedge 1 \leq k \leq K\}$. The landmark is accepted if $\mathrm{var}(R) < \rho$, where var is the variance, and $R = r_k$.

2.3 Visual SLAM

We use a standard implementation of the Extended Kalman Filter (EKF) as basis of the SLAM system [22]. Our method and results, however, are also valid for other SLAM approaches. We will not discuss the full EKF method, rather only the incorporation of the landmark observations.

A landmark i with descriptor $\mathbf{d}_i$ that results from the buffer is classified as either a new landmark, or a previously observed landmark that is already in the map. It concerns a previously observed landmark if the landmark in the database with the most similar descriptor, $\mathbf{d}_j$, fulfills three criteria: 1) Similarity in descriptors: $||\mathbf{d}_i - \mathbf{d}_j|| < \tau_2$, 2) A small best-to-next-best ratio to only match unique landmarks δ_2: $||\mathbf{d}_i - \mathbf{d}_j||/||\mathbf{d}_i - \mathbf{d}_l|| < \delta_2$, where $\mathbf{d}_l$ is the second most similar descriptor in the database, 3) A small distance in the EKF map, measured by the Mahalanobis distance: $\sqrt{(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}_j^{-1}(\mathbf{x}_i - \mathbf{x}_j)} < \eta$, where $\mathbf{S}_j$ is the uncertainty covariance matrix, discussed in the next paragraph.

The landmark is classified as new only if none of the three criteria is fulfilled. For a new landmark, the state matrix and covariance matrix are augmented using the observation, $\mathbf{z}_i$, and the uncertainty covariance matrix, $\mathbf{S}_i$, where $\mathbf{z}_i$ is set to mean(P), and $\mathbf{S}_i$ is determined using the uncertainty of the observation, cov(P), and the uncertainty of the robot's position in the EKF. When a landmark is matched with an existing landmark in the database, $\mathbf{z}_i$ and $\mathbf{S}_i$ are used to update the EKF.

3 Experiments

We performed a number of experiments with a Pioneer II DX robot equipped with a Sony D31 camera to test the use of local symmetry for visual SLAM and to compare it to using standard SIFT. To be able to repeat the experiments, we created a SLAM database. This database contains camera images and odometry information, along with the ground truth position of the robot. The data was recorded in ten different runs, in which the robot drove four laps of approximately 35 meters through an office environment and hallway with an average speed of 0.3 m/s. Camera images of 320 x 240 pixels were stored at 5 Hz. At intervals of one meter, the true location of the robot was logged by hand. This enabled us to quantify the performance of the SLAM estimation.

SIFT and MUST result in different types of interest points, which might have an influence on the best settings for the other parameters in the SLAM system. We therefore optimized the parameters for the visual buffer (N, M, τ_1, δ_1, and ρ) and for the SLAM system (τ_2 and δ_2) for both MUST and SIFT separately. Any differences in performance can therefore be subscribed to the interest-point detectors.
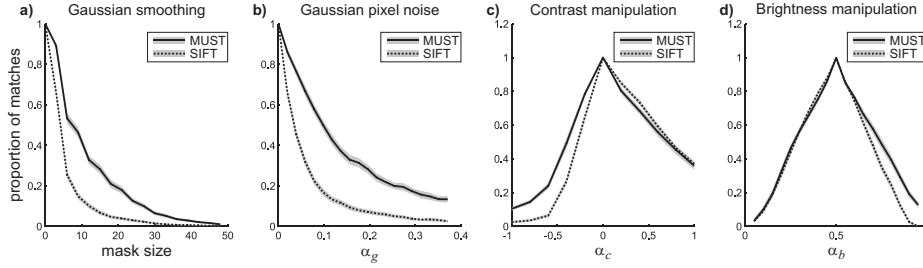
Fig. 4. Robustness to noise and changing light condition. The lines show the mean proportion of matching interest points between the distorted and the original images over 57 runs. The gray areas around the lines display the 98% confidence intervals. Note that these intervals are small, and therefore not very visible.

3.1 Robustness

To test the performance of our model for landmark selection in indoor environments, we took images from one of the runs in our database with intervals of 3 meters to represent the complete environment. To test the noise robustness of MUST, we smoothed the image with a Gaussian kernel, and added Gaussian noise to the pixels. In addition, we manipulated the contrast and brightness of the images to test the robustness to changing light conditions[1]. The reason to not use the dataset in [23] is that it contains images of outdoor scenes, which contain fewer symmetrical objects than indoor environments.

The robustness is measured by the proportion of matching interest points between the original and the manipulated images. Two interest points match when criterion 1 of the visual buffer (see Sect. 2.2) is met, where $\tau_1 = 0.6$ and $\delta_1 = 0.75$. Additionally, the distance between the two points in the image should be less than 3 pixels.

3.2 Repeatability

For good SLAM performance, it is crucial that landmarks added to the map are observable on future encounters. We therefore test the repeatability of the interest points. For every experimental run, we selected three parts of the robot's trajectory, and aligned the sequences of images of the four consecutive laps that are approximately at the same position. The interest points in the images in the first lap are then compared to the images taken at approximately the same position in the later laps.

The average proportion of interest points that are matched in lap 2, 3, and 4 is the measure of repeatability. An interest point matches when criterion 1 of the visual buffer

---

[1] The functions used for the manipulations are: (1) *Gaussian pixel noise*: $I'(x, y) = I(x, y) + X(\alpha_g)$, where $I(x, y) \in [0, 1]$ is the intensity of pixel (x, y), and $X(\alpha_g)$ is a random sample from the normal distribution $N(0, \alpha_g^2)$. (2) *Gaussian smoothing*: $I' = I * G_s$, where G_s is a Gaussian mask of size $s \times s$, with a standard deviation of $\sigma = s/6$. (3) *Contrast manipulation*: $I'(x, y) = I(x, y) + \alpha_c \left(I(x, y) - \bar{I}_{x,y} \right)$, where $\bar{I}_{x,y}$ is the local average in a neighborhood of 21×21 pixels around pixel (x, y). The contrast for $\alpha_c > 0$, and decreases otherwise. (4) *Brightness manipulation*: $'(x, y) = I(x, y)^{\log \alpha_b / \log 0.5}$. For $\alpha > 0.5$, the pixels are brightened, for $\alpha < 0.5$, the pixels are darkened.
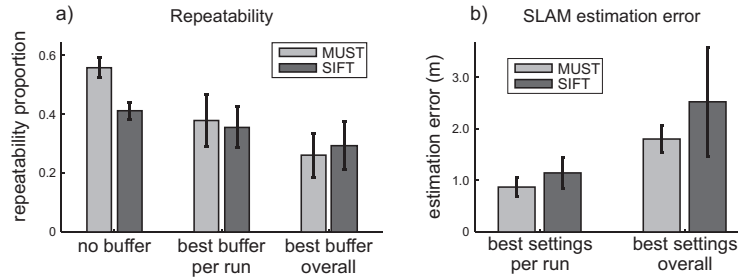
Fig. 5. a) The mean repeatability proportion over the ten runs. The error bars give the 95% confidence intervals. b) The mean estimation error over the ten runs. The 95% confidence intervals are represented by the error bars.

is met (see section 2.2), where where $\tau_1 = 0.6$ and $\delta_1 = 0.75$. Unlike the evaluation of the robustness, we do not put a constraint on the spatial distance between the two observed interest points, since there is variation in the exact position and orientation of the robot. This may cause relatively large displacements of the interest points.

3.3 Visual SLAM Performance

Because we logged the ground-truth position of the robot at intervals of 1 meter, we can quantify the performance of the complete visual SLAM system. For that, we compare the estimated position of the robot made by the EKF with the ground-truth position. The SLAM performance is the average Euclidean distance between the estimated and ground-truth positions in the last of the four laps.

4 Results

4.1 Robustness

Figure 4 shows the robustness results. The proportion of matching interest points between the original and the manipulated images are displayed on the y-axis. In Fig. 4a and 4b, the results for the addition of noise are given. MUST is significantly less affected by the noise than SIFT. The performance of MUST is more than twice that of SIFT, making it much more robust to noise. Figure 4c shows the performance with respect to contrast manipulation. Although the performance of both methods is similar for the contrast enhancement, MUST shows a significantly better performance when the contrast is reduced. Also with enhanced brightness the use of MUST results in considerably better performance (see Fig. 4d). There is no difference with reduced brightness.

4.2 Repeatability

The repeatability results can be appreciated in Fig. 5a. The first two bars show the repeatability results for all detected interest points, in other words when the visual buffer

is not used. The repeatability of the MUST interest points is significantly higher then that of the SIFT points. The two other groups of bars display the results when the buffer is used, where the buffer parameters are optimized both per run and overall with an additional contraint that the buffer does not result in too few (< 50) or too many (> 350) landmarks per run. MUST has a slightly better result per run, whereas overall, SIFT has a somewhat better repeatability. The differences, however, are not significant.

4.3 Visual SLAM Performance

The SLAM performance is displayed in Fig. 5b. The results for the best settings per run show that MUST results in a significantly lower estimation error. Also when the overall best settings are used, MUST performs significantly better than SIFT. This suggests that MUST selects better quality landmarks.

Not only does MUST result in better performance, it also reduces the amount of selected landmarks. MUST selects on average 40 interest points per image, and SIFT 124. This greatly reduces the number of computations in the matching process, which is quadratic in the number if interest points. Since both models are similarly fast in detecting interest points in the images, the full SLAM system is faster using MUST.

5 Discussion

In this paper, we proposed a new interest point detector, MUST, for landmark selection based on local symmetry. The model exploits the presence of symmetrical forms in indoor environments. The interest points selected by our model are less sensitive to noise and changing light conditions and more stable than points selected on the basis of contrast by the SIFT model [4]. The repeatability of all selected interest points is higher for our model than for SIFT. Most importantly, the use of symmetry results in better visual SLAM performance. This shows that exploiting symmetry results in valuable and robust landmarks for the representation of the environment. Moreover, MUST results in less interest points, which improves the processing time of the SLAM system.

Although the visual buffer successfully reduces the number of landmarks in the map, its effectiveness can be further enhanced, judging from the fact that the repeatability drops when the visual buffer is used.

Symmetry detection can results in the selection of complete objects. This could result in semantically more meaningful landmarks and maps of the environment, which could be exploited, for instance, in human-robot interaction.

We can conclude that local symmetry provides robust and valuable landmarks, that results in improved visual SLAM performance.

References

1. Thrun, S., Burgard, W., Fox, D.: Probabilistic Robotics. The MIT Press, Cambridge, Massachusetts (2005)
2. Frintrop, S., Jensfelt, P.: Attentional landmarks and active gaze control for visual slam. IEEE Transactions on Robotics **24**(5) (2008)

3. Davison, A.J., Reid, I.D., Molton, N.D., Stasse, O.: Monoslam: Real-time single camera slam. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(6) (2007) 1–16

4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2) (2004) 91–110

5. Se, S., Lowe, D.G., Little, J.: Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. International Journal of Robotics Research **21**(8) (2002) 735–758

6. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: ninth European Conference on Computer Vision (ECCV), Graz, Austria (2006)

7. Murillo, A.C., Guerrero, J.J., Sagues, C.: Surf features for efficient robot localization with omnidirectional images. In: IEEE International Conference on Robotics and Automation (ICRA), Rome, Italy (2007) 3901–3907

8. Davison, A.J., Murray, D.: Simultaneous localization and map-building using active vision. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7) (2002) 865–880

9. Mozos, O.M., Gil, A., Ballesta, M., Reinoso, O.: Interest point detectors for visual slam. In Borrajo, D., Castillo, L., Corchado, J.M., eds.: Lecture Notes in Computer Science. Volume 4788. Springer-Verlag (2008) 170–179

10. Moreels, P., Perona, P.: Evaluation of features detectors and descriptors based on 3d objects. International Journal of Computer Vision **73**(3) (2007) 263–284

11. Palmer, S.E., Hemenway, K.: Orientation and symmetry: Effects of multiple, rotational, and near symmetries. Journal of Experimental Psychology: Human Perception and Performance **4**(4) (1978) 691–702

12. Kootstra, G., Nederveen, A., de Boer, B.: Paying attention to symmetry. In Everingham, M., Needham, C., Fraile, R., eds.: British Machine Vision Conference (BMVC2008), Leeds, UK (2008) 1115–1125

13. Kootstra, G., Schomaker, L.R.: Prediction of human eye fixations using symmetry. In: Cognitive Science Conference (CogSci), Amsterdam, NL (2009)

14. Driver, J., Baylis, G.C., Rafal, R.D.: Preserved figure-ground segregation and symmetry perception in visual neglect. Nature **360** (1992) 73–75

15. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(11) (1998) 1254–1259

16. Marola, G.: Using symmetry for detecting and locating objects in a picture. Computer Vision, Graphics, and Image Processing **46** (1989) 179–195

17. Backer, G., Mertsching, B., Bollmann, M.: Data- and model-driven gaze control for an active-vision system. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(12) (2001) 1415–1429

18. Sela, G., Levine, M.D.: Real-time attention for robotic vision. Real-Time Imaging **3** (1997) 173–194

19. Heidemann, G.: Focus-of-attention from local color symmetries. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(7) (2004) 817–830

20. Reisfeld, D., Wolfson, H., Yeshurun, Y.: Context free attentional operators: The generalized symmetry transform. International Journal of Computer Vision **14** (1995) 119–130

21. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(8) (2003) 959–973

22. Durrant-Whyte, H., Bailey, T.: Simultaneous localization and mapping: Part i. Simultaneous Localization and Mapping: Part I. **13**(2) (2006) 99–108

23. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. International Journal of Computer Vision **65**(1/2) (2005) 43–72