# Evaluation of interest point detectors for Visual SLAM

Mónica Ballesta, Arturo Gil, Óscar Reinoso
Systems Engineering Department
Miguel Hernández University
03202 Elche (Alicante), SPAIN
Email: {mballesta, arturo.gil,o.reinoso}@umh.es

Óscar Martínez Mozos
University of Freiburg
Department of Computer Science
79110 Freiburg, Germany
Email: omartine@informatik.uni-freiburg.de

***Abstract:*** **In this paper we present several interest points detectors and we analyze their suitability when used as landmark extractors for vision-based simultaneous localization and mapping. For this purpose, we evaluate the detectors according to their repeatability under changes in viewpoint and scale. These are the desired requirements for visual landmarks. Several experiments were carried out using sequence of images captured with high precision. The sequences represent planar objects as well as 3D scenes.**

**Key-Words: Interest-Point detector, Visual SLAM, Visual Landmarks, Mobile Robots**

## 1. INTRODUCTION

Acquiring maps of the environment is a fundamental task for autonomous mobile robots, since the maps are required for different higher level tasks. As a result, the problem of simultaneous localization and mapping (SLAM) has received significant attention. Typical approaches use range sensors to build maps in two and three dimensions (see, for example, [1, 2, 3] [4, 5, 6]). In recent years there is an increasing interest on using cameras as sensors. Such approach is sometimes denoted as visual SLAM. Cameras offer higher amount of information and are less expensive than lasers. Moreover, they can provide 3D information when stereo systems are used.

Usual approaches using vision apply a feature-based SLAM, in which visual features are used as landmarks. The main issue when performing visual SLAM is how to select suitable features on the images to be used as reliable landmarks. When the map to construct has three dimensions, the landmarks must be additionally robust to changes in the scale and viewpoint. Different vision features has been used for mapping and localization using monocular or stereo vision, as for example, lines [7], region of interest [8]; and interest points as SIFT [9, 10, 11], Harris corner detector [12, 13] or SURF [14]. The interest points detectors have received most of the attention in visual SLAM. The points detected are typically invariant under rotation, translation, scale and only partially invariant under changes in viewpoint. These theoretical properties make them suitable for being used as visual landmarks. In practice, however, the detected points are not stable, and the the matching between different views becomes difficult. Some solutions have been applied to solve this problem, as mixing several methods in one detector [15] or tracking the points during several frames to keep the stability [16, 10]. However, the question of which interest point detector is more suitable for visual SLAM is still open.

In this paper we present several evaluations of different point detectors that are typically used in visual SLAM. The extracted points used as landmarks should be robust under scale and viewpoint changes. These requirements are necessary for SLAM, since the robot must be able to detect and associate new landmarks to previous ones, even if they are observed from significantly different viewpoints. Under these conditions we analyze the repeatability of the points in consecutive images and the probability of been detected in future ones.

The rest of the paper is organized as follows. After discussing some related work in Section 2., we present in different interest point detectors Section 3.. Section 4. introduces the evaluation methods used in this work. Several experiments are presented in Section 5.. We finally conclude in Section 6..

## 2. RELATED WORK

Visual SLAM has been an interesting topic in mobile robotics for the last years. Different methods has been used to extract visual landmarks. Lemaire and Lacroix [7] use segments as landmarks together with and EKF-based SLAM approach. Frintrop *et al.* [8] extract regions of interest (ROI) using the attentional system VOCUS. Several authors use SIFT features as landmarks in the 3D space [9, 16]. Little *et al.* [17]. Gil *et al.* [10] additionally track the SIFT features to keep the most robust ones; and Valls Miró *et al.* [11] use SIFT to map medium sized environments. Harris corner detectors has also been used as landmarks for monocular SLAM (Davison and Murray [12]) or in Autonomous Blimps (Hygounenc *et al.* [13]). Finally, Murillo *et al.* [14] present a localization method using SURF features.

In the context of matching and recognition, many authors have presented their works evaluating several interest point detectors. The work presented by Mikolajczyk and Schmid [18], uses different detectors to extract affine invariant regions, but only focuses on the comparison of different description methods. In [19], a collection of detectors is evaluated. The criteria used measures the quality of these features for tasks like image matching, object recognition and 3D reconstruction. However they do not take into account the repeatability in successive frames of a sequence. In contrast to the previous works we evaluate the different interest point detectors under the particular conditions of visual SLAM.

## 3. INTEREST POINT DETECTORS

Along this paper we suppose that a mobile robot is used for constructing the map of the environment. The robot is equipped with a camera used to acquire images. Interest points are then extracted from these images and used as landmarks. We also suppose that the height of the camera on the robot is fixed as well as its orientation. This is the typical configuration in visual SLAM systems. Additionally, we assume that visual landmarks are static, i. e. they do not change their position or orientation during the experiments. According to the previous criterion, we following present five different interest point detectors used to extract visual landmarks.

### 3.1. Harris Corner Detector

The Harris Corner Detector [20] is probably the most widely interest point detector used due to its strong invariance to scale, rotation and illumination variations, as well as image noise. The detector is based on the matrix $C(x, y)$ which is computed over a $p$x$p$ patch for each interest point at position $(x, y)$ as:

$$C(x,y) = \begin{pmatrix} \sum I_x^2 & \sum I_x I_y \\ \sum I_x I_y & \sum I_y^2 \end{pmatrix} \quad , \tag{3.1}$$

where $I_x$,$I_y$ are the image gradients in horizontal and vertical direction. Let $\lambda_1$ and $\lambda_2$ be the eigenvalues of the matrix $C(x, y)$, we define the auto-correlation function $R$ as:

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \ . \tag{3.2}$$

This function will be sharply peaked if both of the eigenvalues are high. This means that shifts in any direction will produce a significant increase, indicating that it is a corner. A typical value for $k$ is $0.04$ [12].

### 3.2. Harris-Laplace

The interest points extracted by the Harris-Laplace detector [21] are invariant to rotation and scale. These points are detected by a scale adapted Harris function and selected in scale-space by the Laplacian operator. The selected scale determines the size of the support region.

### 3.3. SIFT

The Scale-Invariant Feature Transform (SIFT) is an algorithm that detects distinctive keypoints from images and computes a descriptor for them. This algorithm was initially presented by Lowe [22] and used in object recognition tasks. The interest points extracted are said to be invariant to image scale, rotation, and partially invariant to changes in viewpoint and illumination. SIFT features are located at maxima and minima of a difference of Gaussians (DoG) function applied in scale space. They can be computed by building an image pyramid with resampling between each level [23]. In this work, we only use the detected points and we discard the descriptors.

### 3.4. SURF

Speeded Up Robust Features (SURF) is a scale and rotation invariant detector and descriptor which was recently presented by Bay *et al.* [24]. This detector is based on the Hessian matrix because of its accuracy and low computational time. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images. According to [24], this algorithm outperforms existing methods with respect to repeatability, robustness and distinctiveness of the descriptors. As with SIFT features, we concentrate only on the detected points and we discard the descriptors.

### 3.5. SUSAN

SUSAN (Smallest Univalue Segment Assimilating Nucleus) is an approach to low level image processing [25]. The SUSAN principle is implemented using digital approximation of circular masks. If the brightness of each pixel within a mask is compared with the brightness of that mask's nucleus, then an area of the mask can be defined which has the same brightness as the nucleus. SUSAN has been traditionally used for object recognition.

## 4. EVALUATION METHODS

To evaluate the previous methods we use sequences of images representing the same scene under different scales and viewpoints. In this section we explain how these sequences were evaluated. We first introduce the tracking methods used to follow the interest points in each frame of the sequences. We then describe the measurements used to study the repeatability and robustness of each method under changes in scale an viewpoint. In this work we do not study the invariance under changes in illumination.

### 4.1. Tracking

For each image in a sequence, we first extract the interest points using the methods explained in Section 3. In order to track each point in successive 2D images we try to match the interest points using the homography matrix for each pair of consecutive images as follows [26]. Given a point $Y$ in 3D space, we assume that this point projects at position $y_1 = P_1 Y$ in image $I_1$ and at position $y_i = P_i Y$ in image $I_i$, with projection matrices $P_1$ and $P_i$. If we suppose that the point $Y$ is detected in both images, then

$$y_i = H_{1i} \times y_1, \text{ with } H_{1i} = P_i P_1^{-1} \ . \qquad (4.3)$$

The homography matrix $H_{1i}$ can be computed by selecting manually four correspondences of coplanar points between images 1 and $i$. Given a detected point in one image, we predict its position in the consecutive image using the homography matrix. If the predicted position lies at a distance below 2 pixels from an interest point detected in the second image, then we consider that the interest point is successfully tracked. If no interest point lies in the neighborhood of the predicted point, then the tracking of the point is lost. This method has been applied to sequences of images containing planar objects, since the computation of the homography matrix can only be made for coplanar points in the space.

In the case of images with 3D scenarios, the tracking has been implemented by using the fundamental matrix. This 3x3 matrix with rank 2 relates the corresponding points between two stereo images and can be implemented with at least 7 corresponding points. Given a point $m_1$ from image $I_1$, the fundamental matrix F computes the epipolar line of the second image $I_2$ where the corresponding point $m'_1$ must lie. The epipolar line is computed as $l' = Fm_1$ (see Figure 2). In consequence, two corresponding points will satisfy the following equation

$$m_i'^T F m_i = 0 \ . \qquad (4.4)$$

In our work, we have computed a set of previous correspondences using the fundamental matrix with the 7 points method. This points have been chosen manually. In order to select candidate points of a point in the previous image, we have establish a search window of 10x10 pixels around this point in the consecutive image. Then the correspondent point is select among the candidates as the point which has the smallest distance to the epipolar line. This distance has been computed as follows

$$d(m'_i, Fm_i) = \frac{|m_i'^T F m_i|}{\sqrt{(Fm_i)_1^2 + (Fm_i)_2^2}} \ . \qquad (4.5)$$

As a result of this previous step we have an array of correspondent points between both images. In a second step, we have use these previous correspondences as the input for the computation of a second fundamental matrix which uses the RANSAC algorithm as in [27]. This will be a more accurately computed matrix than allows us to obtain the definitive correspondences. We have used the


0–1 1–2 2–3 3–4 4–5 5–6

Figure 1: Sequence of images with persistent points (red), lost points (blue) and points detected (white).

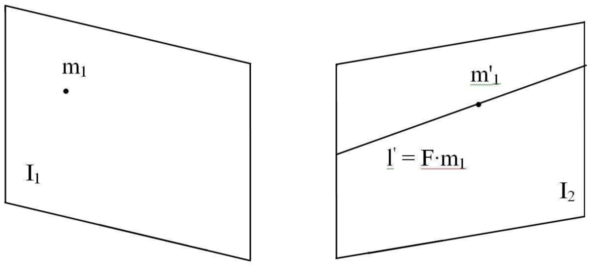


Figure 2: The point $m_1'$ is the corresponding point of $m_1$ in the image $I_2$. This point lies in the epipolar line $l'$ computed with the fundamental matrix $F$.

same search window as before and the incorrect candidates have been also rejected using the distance of equation 4.5.

The usage of these tracking methods that are only based on geometric measures avoids making any kind of description of this points, which allows us to focus our work on the detection stage.

An example of a tracking using one of these methods is shown in Figure 1 in which the interest points were extracted with the Harris detector (white points). In this sequence, the red points in the last image indicate points that could be tracked along the whole sequence. The blue points are those ones that have been lost from the previous image. A point that is lost, even only once, is rejected by our tracking algorithm since we have considered that this kind of points are not stable enough for our purpose.

## 4.2. Evaluation Measurements

As explained in Section 3., we want to evaluate the detectors according to the SLAM requirements. In this sense, we have follow a repeatability criterion which means that the detection is independent of changes in the imaging conditions, i.e. scale and viewpoint. Applying our tracking method we first define the survival ratio $S_i$ in the frame $i$ as:

$$S_i = \frac{\mathrm{np}_i}{\mathrm{np}_0} \cdot 100 \ , \tag{4.6}$$

where $\mathrm{np}_i$ and $\mathrm{np}_0$ are the number of points detected in the frame $i$ and the first frame respectively. A perfect detector would detect the same points in the first and the last frame, i. e. $S_i = 100\%$ for every frame. However, as we will see in the experiments, we normally observe decreasing tendency in $S_i$, meaning that some of the points observed in the first frame are lost in subsequent frames.

When the robot explores the environment, it is desirable to extract visual landmarks that are stable and can be detected in a number of $p$ consecutive frames [17, 10]. As a result, the number of landmarks in the map is reduced and also the complexity of the SLAM problem. However, setting $p$ poses a problem: if $p$ is low, a high number of spurious points will be integrated in the map. If $p$ is high, the number of landmarks in the map will be too low. For example, when the robot turns, the landmarks disappear rapidly from the camera field of view and will not be integrated in the map if $p$ is high. Taking into account this requirement we analyze for how many frames we should track a landmark before integrating it in the map. We use the following conditional probability:

$$P(t_{f_a}|t_{f_b}) = \frac{t_{f_a}}{t_{f_b}}, \tag{4.7}$$

where $t_{f_i}$ is the number of points tracked until frame $f_i$.

This value represents the probability of an interest point to be tracked until frame $f_a$ given that it was tracked until frame $f_b$. This value ranges between $0$ and $1$. It is $0$ when all points tracked until $f_b$ are lost in frame $f_a$, and $1$ if both frames $f_a$ and $f_b$ contains the same tracked points.

Expression (4.7) gives a prediction of the survival of an interest point in future frames if the movement of the robot maintains similar. This expression can be used to estimate the number of frames $p$ a landmark has to be tracked before it is incorporate in the map (Section 5.).

## 5. EXPERIMENTS

In order to evaluate the different interest point detectors, we captured 12 sequences of viewpoint changing images each containing 21 images. For each image we increased the angle in $2.5$ degrees. Additionally we captured 14 sequences of images with scale changes each containing 12 images. In this last case the camera moved 0.1 meters in each image. The sequences contain images of planar objects (as posters) and 3D scenes. Examples of both types are shown in Figure 3 and Figure 4 respectively.

All images were captured using a STH-MDCS2 stereo head from Videre Design. Only one of the stereo images was used at each time to form the sequences. The stereo head was mounted on a robotic arm to achieve constant variations of viewing angle and distance change. Finally, the images were captured at different resolutions (320x240, 640x480 and 1280x960), so that the set of images could be as much representative as possible.

In this paper, Figures [ 5, 6] are referred to 2D images and Figures [ 7, 8] to 3D images. For those images two different ways of comparing the results have been implemented.

In a first experiment we analyze the repeatability of the different detectors in the sequences with different viewpoints. In SLAM it is important that landmarks detected with a certain angle and distance are also detected from different ones. This comes from the fact that a mobile robot will see the same point in the scene from different poses in the environment. For this purpose we use at the same time all the sequences captured and we calculate Expression (4.6) taking into account the interested points of all images in each case. This is shown in the left Figures (Figures (a)), where the Harris corner detector seems to be more stable. In 2D images a $30\%$ of the points detected with Harris survive all the sequence in images with viewpoint changes (Figure 5) and with scale transformation (Figure 6). In the case of 3D images, this result is even better achieving around $50\%$ (see Figures 7 and 8).

The figures of the right present a different way of comparing these detectors. In this case,the plots show the probability that a point is found in the last frame given that it was tracked until the frame $i$ as shown in Expression (4.7). Again the Harris corner detector shows the best results with scale and viewpoint changes in all the images(2D and 3D). For example, in Figure 7(b) is shown that a Harris-point that is detected in the first 6 frames has a probability of 0.7 of being tracked until frame 20.

Harris detector is the best one in all cases, but comparing the results between 2D and 3D, it is shown that Harris performs even better with 3D images (Figures 7, 8).In this figures, the Harris corner detector has also a better result in comparison with the other detectors. In Figure 7(a) a $55\%$ of the Harris-points is tracked until the last frame, whereas the other detectors only achieve around a $20\%$.

With respect to the rest of the detectors, the figures show that Harris detector is followed by Harris Laplace, SIFT and SURF which have a similar behavior. Particularly, SURF performs better with 3D images (Figures [ 7, 8]) as well as with scale changes in general. For instance, in Figure 6 a point detected by SURF which is followed 6 frames has a probability of 0.5 of being detected until the last frame. In this case, SURF has the second best behavior. SIFT is the second best detector in 2D images with viewpoint changes (Figures 5(a), 5(b)) but it has worse results in other cases. Finally, SUSAN is the one that performs worse in all cases.

Tables 1 and 2 present the number of interest points detected in the first image and the number of points that were tracked until the last frame. It can be clearly seen that the number of points detected differs when using different methods. This stems from the fact that we are using an heterogeneous image database and it is not possible to adjust each of the detectors in a way that the number of detected points is the same for all the methods. For instance, the parameters for each of the five methods can be adjusted in a way that the number of points detected in a single image would be equal. However, the same parameters applied to a different image would result in differing number of points detected. In consequence, the results presented here are normalized to the number of points that appear in the first frame, so that they can be compared.


Figure 3: The top sequence shows images of a poster from different viewpoints. The bootom sequence shows the same poster with changes in scale.

Table 1: Number of points detected in the first and last image of each sequence of 2D images

| **Images 2D** | | | | | |
|---|---|---|---|---|---|
| **Changes in Viewpoint** | Harris | Harris Laplace | SUSAN | SIFT | SURF |
| Number of points detected in the first image | 1051 | 1216 | 1472 | 1830 | 5438 |
| Number of points tracked to the last image | 321 | 170 | 18 | 407 | 853 |
| **Changes in Scale** | Harris | Harris Laplace | SUSAN | SIFT | SURF |
| Number of points detected in the first image | 4507 | 4431 | 5004 | 6330 | 20335 |
| Number of points tracked to the last image | 1237 | 606 | 392 | 796 | 3458 |

## 6. CONCLUSION

In this paper we presented an evaluation of different interest point detectors. We focused on the use of interest points in vision-based SLAM. For this purpose we analyzed each detector according to the properties desired for visual landmarks: repeatability and accuracy. The results of the experiments showed the behavior of five different detectors under changes in viewpoint and scale. We believe that this information will be useful when selecting an interest point detector as visual landmark extractor in SLAM.

## Acknowledgments

This research has been supported by the spanish government through project DPI2004-07433-C02-01 (Ministerio de Educación y Ciencia. Título: HERRAMIENTAS DE TELEOPERACIÓN COLABORATIVA. APLICACIÓN AL CONTROL COOPERATIVO DE ROBOTS), and project PCT-G54016977-2005 (FUNDACIÓN QUORUM: PARQUE CIENTÍFICO Y EMPRESARIAL DE LA UNIVERSIDAD MIGUEL HERNÁNDEZ. Título: ROBOTS COOPERATIVOS PARA LA VIGILANCIA E INSPECCIÓN DE EDIFICIOS E INSTALACIONES INDUSTRIALES).

## References

[1] G. Grisetti, C. Stachniss, and W. Burgard. Improved techniques for grid mapping with rao-blackwellized particle filters. *IEEE Transactions on Robotics*, 23(1), 2007.

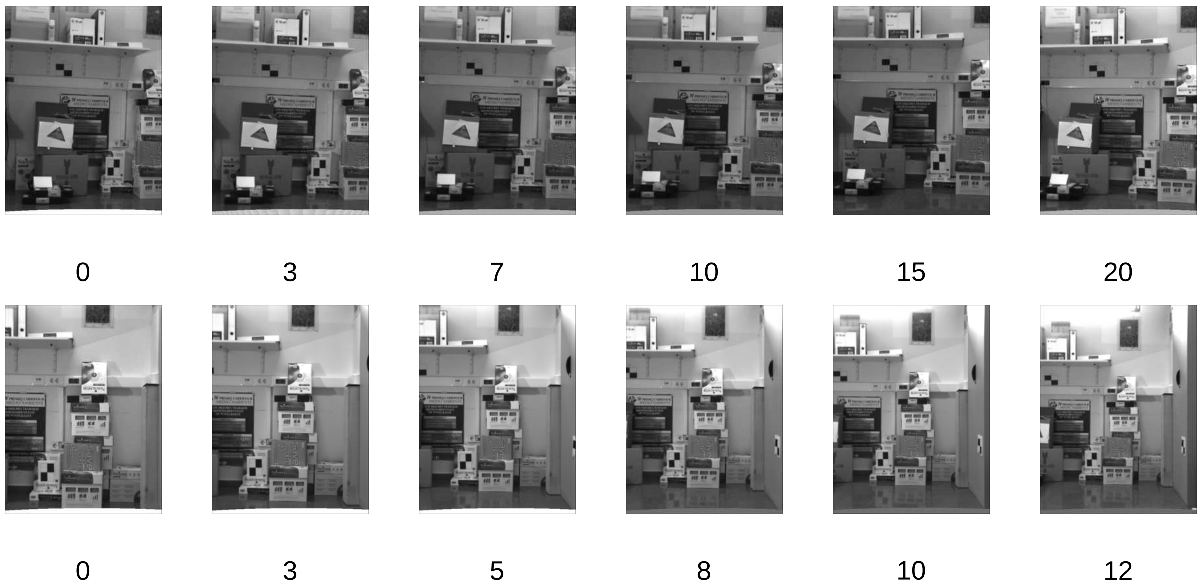


Figure 4: The top sequence shows images of a 3D scene from different viewpoints. The bottom sequence shows a similar scene with changes in scale.

Table 2: Number of points detected in the first and last image of each sequence of 3D images

| **Images 3D** | | | | | |
|---|---|---|---|---|---|
| **Changes in Viewpoint** | Harris | Harris Laplace | SUSAN | SIFT | SURF |
| Number of points detected in the first image | 1013 | 1372 | 1495 | 1774 | 4934 |
| Number of points tracked to the last image | 548 | 276 | 159 | 356 | 1132 |
| **Changes in Scale** | Harris | Harris Laplace | SUSAN | SIFT | SURF |
| Number of points detected in the first image | 963 | 1254 | 1417 | 1668 | 4661 |
| Number of points tracked to the last image | 451 | 313 | 220 | 329 | 1212 |


[2] D. Hähnel, W. Burgard, D. Fox, and S. Thrun. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. In *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, Las Vegas, NV, USA, 2003.

[3] J.J. Leonard and H.F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(4), 1991.

[4] P. Biber, H. Andreasson, T. Duckett, and A. Schilling. 3d modelling of indoor environments by a mobile robot with a laser scanner and panoramic camera. In *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, 2004.

[5] R. Eustice, H. Singh, and J.J. Leonard. Exactly sparse delayed-state filters. In *IEEE Int. Conf. on Robotics & Automation*, 2005.

[6] R. Triebel and W. Burgard. Improving simultaneous mapping and localization in 3d using global constraints. In *National Conference on Artificial Intelligence*, 2005.

[7] Thomas Lemaire and Simon Lacroix. Monocular-vision based SLAM using line segments. In *IEEE Int. Conf. on Robotics & Automation*, 2007.

[8] S. Frintrop, P. Jensfelt, and H. I. Christensen. Attentional landmark selection for visual slam. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2006.

[9] J. Little, S. Se, and D.G. Lowe. Vision-based mobile robot localization and mapping using scale-

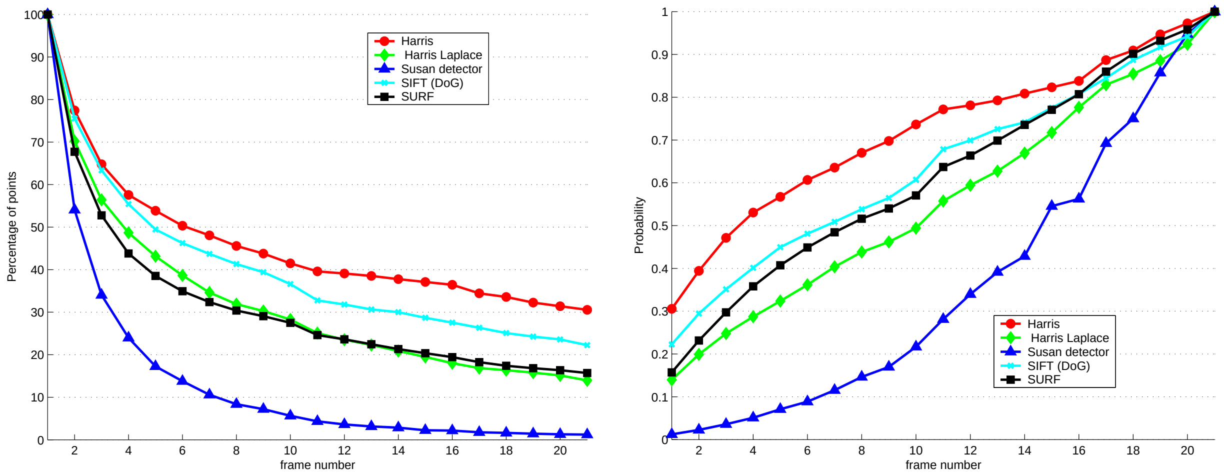


Figure 5: Images 2D with viewpoint transformation. Fig. 5(a) shows the survival ratio for each of the frames in the sequence. Fig. 5(b) shows the probability of a point being detected in the last frame given that it was detected in the frame i.

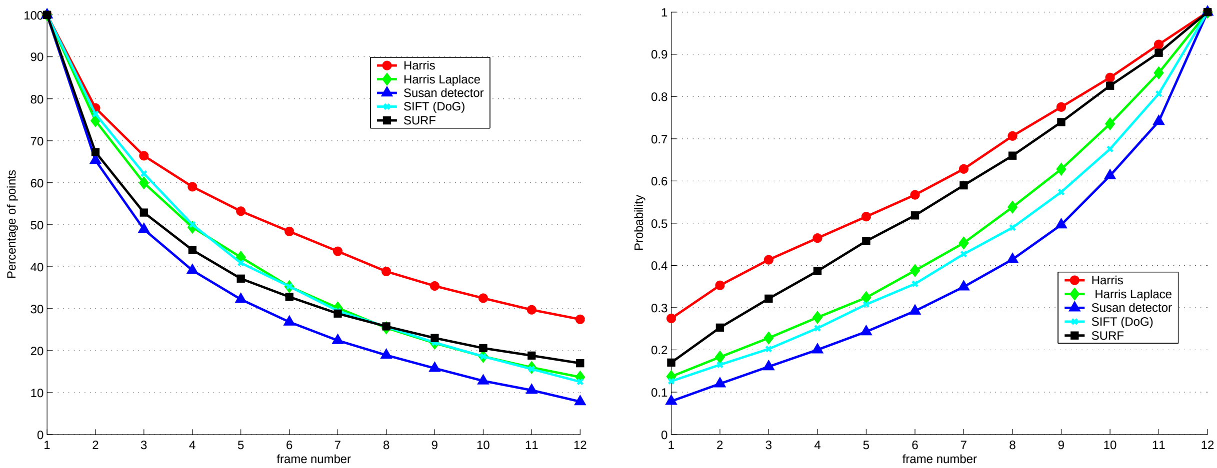


Figure 6: Images 2D with scale transformation. Fig. 6(a) shows the survival ratio for each of the frames in the sequence. Fig. 6(b) shows the probability of a point being detected in the last frame given that it was detected in the frame i.


invariant features. In *IEEE Int. Conf. on Robotics & Automation*, 2001.

[10] A. Gil, O. Reinoso, W. Burgard, C. Stachniss, and O. Martínez Mozos. Improving data association in rao-blackwellized visual SLAM. In *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, 2006.

[11] J. Valls Miro, W. Zhou, and G. Dissanayake. Towards vision based navigation in large indoor environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, 2006.

[12] Andrew J. Davison and David W. Murray. Simultaneous localisation and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002.

[13] Emmanuel Hygounenc, Il-Kyun Jung, Philippe Souères, and Simon Lacroix. The autonomous blimp project of laas-cnrs: Achievements in flight control and terrain mapping. *International Journal of Robotics Research*, 23(4–5), 2004.

[14] A. C. Murillo, J. J. Guerrero, and C. Sagüés. Surf features for efficient robot localization with omnidirectional images. In *IEEE Int. Conf. on Robotics & Automation*, 2007.

[15] Patric Jensfelt, Danica Kragic, John Folkesson, and Mårten Björkman. A framework for vision based

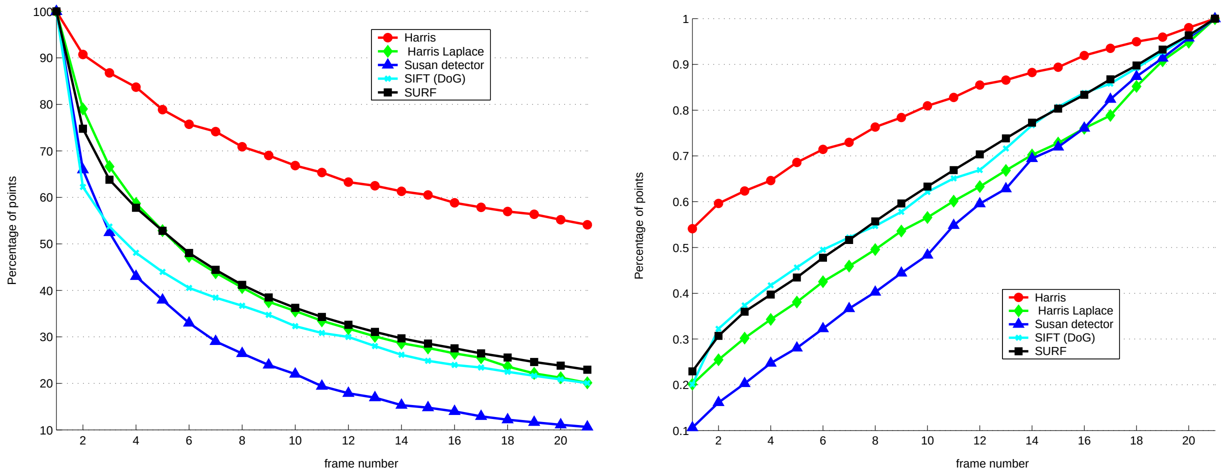


Figure 7: Images 3D with viewpoint transformation. Fig. 7(a) shows the survival ratio for each of the frames in the sequence. Fig. 7(b) shows the probability of a point being detected in the last frame given that it was detected in the frame i.

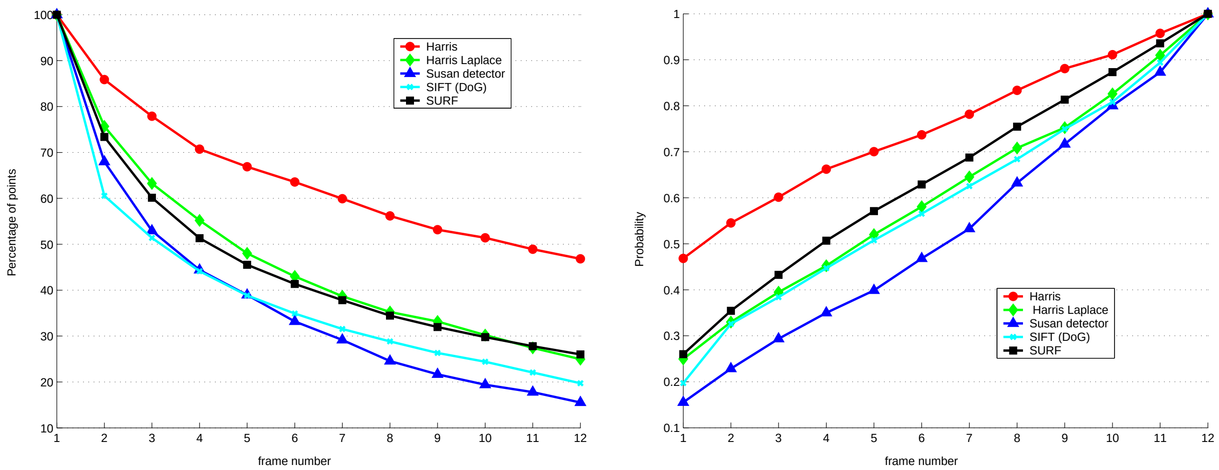


Figure 8: Images 3D with scale transformation. Fig. 8(a) shows the survival ratio for each of the frames in the sequence. Fig. 8(b) shows the probability of a point being detected in the last frame given that it was detected in the frame i.

bearing only 3D SLAM. In *IEEE Int. Conf. on Robotics & Automation*, 2006.

[16] Stephen Se, David G. Lowe, and Jim Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *IEEE Int. Conf. on Robotics & Automation*, 2001.

[17] J. Little, S. Se, and D.G. Lowe. Global localization using distinctive visual features. In *IEEE/RSJ Int. Conf. on Intelligent Robots & Systems*, 2002.

[18] K. Mikolajczyk and C Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 2005.

[19] C. Schmid, R. Mohr, and C. Bauckhage. Evaluaton of interest point detectors. *International Journal of computer Vision*, 37(2), 2000.

[20] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1998.

[21] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *Int. Conf. on Computer Vision*, 2001.

[22] D.G. Lowe. Object recognition from local scale-invariant features. In *Int. Conf. on Computer Vision*, 1999.

[23] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of computer Vision*, 2(60), 2004.

[24] H. Bay, T. Tuytelaars, and L. Van Gool. Object recognition from local scale-invariant features. In *European Conference on Computer Vision*, 2006.

[25] S.M. Smith. A new class of corner finder. In *British Machine Vision Conference*, 1992.

[26] Gy. Dork and C. Schmid. Selection of scale invariant neighborhoods for object class recognition. In *Int. Conf. on Computer Vision*, 2003.

[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. In *Artificial Intelligence*, volume 78 pp. 87-119, 1995.