

Adaptive Dominant Point Detector for Visual Landmark Description*

Esther Antúnez¹, Rebeca Marfil², Antonio Bandera², Ricardo Vázquez-Martín²
and Walter G. Kropatsch¹

¹ PRIP, Vienna University of Technology, Austria
{eortiz,krw}@prip.tuwien.ac.at

² ISIS, Departamento de Tecnología Electrónica
Universidad de Málaga, Spain
{rebeca,ajbandera,rvmartin}@uma.es

Abstract

This paper describes a visual landmark detection and description system for mobile robotic environment mapping. At the detection stage, the paper employs a perceptual-based hierarchical algorithm which has been previously used for segmentation of natural images. At the description stage, visual landmarks are characterized by a kernel-based representation and by the set of dominant points extracted from their outer boundaries. These dominant points allow to represent the landmark region by a set of stable points which can be located –by its position and uncertainty– on the 3D real world. Although all system modules are briefly described and evaluated, this paper is specifically focused on the dominant point detector.

1 Introduction

Reliable navigation is an essential component of an autonomous mobile robot which typically implies representation of the information perceived by external sensors into an internal navigation map. It is interesting that this map can be built with distinguished natural landmarks that the robot acquires from the environment without human supervision. In this framework, landmarks can be defined as “*distinct features that a vehicle can recognize reliably from its sensor observations*” [9]. Recognizable landmarks are essential since they will be used as reference marks to identify world locations. To detect these landmarks, vision systems have been considered in this last years as an attractive alternative to the active ranging devices [18]. These systems are passive and of high resolution, providing a huge amount of features (color, texture or shape) that can permit to disambiguate landmarks for subsequent data association purposes.

This paper describes a visual landmark detection and description system for environment mapping. In this system, distinguished regions are extracted from the input image by a perception-based grouping mechanism. The central item of the perceptual grouping approach for image segmentation is an irregular pyramid: the Bounded Irregular Pyramid (BIP) [11, 10]. However, the BIP presents an inherent problem which can damage its application on this framework: it provides an image segmentation

*This work has been partially supported by the Austrian Science Fond under grants P18716-N13 and S9103-N13 and by the Spanish Junta de Andalucía project P07-TIC-03106.

which varies when this image is shifted slightly [11]. To avoid this shift–variance problem, the BIP has been modified, changing the decimation process used to build its internal structure. From the segmentation regions provided by this perceptual–based grouping algorithm, the set of image regions which satisfies certain criteria is chosen. These visual features are then characterized by a kernel–based representation and by the set of dominant points detected over its outer contour. This paper is mainly focused on this last issue.

The rest of the paper is organized as follows: after presenting an overview of the proposed system in Section 2, Section 3 describes the employed dominant point detector. Experimental results in Section 4 show the performance of the different system modules, providing comparative studies where they have been evaluated. Finally, in Section 5, we draw the main conclusions and outline future research directions.

2 Overview of the Proposed System

The visual landmark detection and description system consists of three modules or stages. First, a hierarchical perceptual grouping algorithm is applied to perform a domain–independent segmentation of the image pixels into regions. Thus, although the final obtained regions do not always correspond to the natural image objects, it provides a mid–level segmentation which is more coherent with the human–based image decomposition. Then, the set of regions which satisfies certain rules are selected as landmarks. These rules do not depend on the environment or application, and they impose to the obtained landmarks properties like good contour closure and continuity or high contrast with respect to its surrounding background [10]. Besides, the shape of these regions is adapted to real items of the scene. Therefore, continuous geometric transformations preserve topology pixels from a single connected item are transformed to a single connected item. Thus, after a geometric change locally approximated by an affine transform, homography or even continuous non–linear warping, a perceptually homogeneous region will be in the transformed set of regions (see Fig. 1). Obtained landmarks are characterized by the 3D position of a set of dominant points detected over its outer contour. This allows to locate the region on the 3D real world. Finally, a kernel–based representation [4] is employed to describe its internal colour distribution.

The modules involved in the solution of the visual landmark detection and description system are briefly described below:

- *Hierarchical perceptual grouping algorithm.* The hierarchical perceptual grouping algorithm performs the segmentation of the input image using two consecutive stages. The first stage employs a colour distance to group the image pixels into a set of regions whose spatial distribution is physically representative of the image content. Then, the second stage groups the set of homogeneous regions into a smaller set of regions taking into account not only the internal visual coherence of the obtained regions but also the external relationships among them. This scheme has been successfully employed to segment natural images (see [10] for further details). In this approach, the segmentation is accomplished using an irregular pyramid: the Bounded Irregular Pyramid (BIP). The BIP is a mixture of regular and irregular pyramids, i.e. the BIP combines a regular structure with an irregular simple graph. The regular decimation is applied in the homogeneous parts of the image, meanwhile the heterogeneous parts are decimated using a classical irregular process [11]. Currently, the structure of the BIP has been modified in order to improve the mixture of the regular and irregular decimation processes. This new pyramid



Figure 1: Regions generated by the proposed detector on two images taken from different viewpoints. Regions are represented by ellipses which have their same first and second moments. Then, ellipses show the original detection size.

allows to merge regular and irregular pyramid nodes, linking them to the same parent node. Table 1 presents the obtained results from the evaluation of several decimation schemes using the shift-variance (SV) test [11]. It must be noted that the smaller the value of this parameter (SV), the better the segmentation result should be. Experimental results show that the modified BIP decimation scheme is robust against slightly shiftings of the input image.

The hierarchical perceptual grouping approach provides a partition of the input image into a set of regions. Among these regions, the approach selects those which satisfy certain conditions. Thus, selected regions cannot be located at the image border in order to avoid errors due to partial occlusions. They must also exhibit a relatively high color contrast with respect to its surrounding regions. Specifically, the conditions imposed to be a landmark are the following:

- The bounding box of a selected region must not be located in an image border. The bounding box of a region is the minimum box which encloses it.
 - The contrast between the color of a landmark and its surrounding regions will be higher than a threshold value.
- *Landmark appearance description.* Vision can be useful to avoid data association failures allowing landmarks to be characterized by a robust descriptor, i.e. a descriptor which will be invariant to illumination and viewpoint changes. In our case, this descriptor is estimated from an image region whose shape changes as a function of the image acquisition viewpoint. This

	MIS [13]	D3P [6]	MIES [5]	BIP [11]	Modified BIP
SV_{min}	39.9	31.8	23.7	25.6	19.5
SV_{ave}	59.8	49.1	44.1	73.8	43.7
SV_{max}	101.1	75.3	77.2	145.0	73.2

Table 1: Shift Variance values for different decimation processes. Average values have been obtained from 30 color images from Waterloo and Coil 100 databases (All these images have been resized to 128x128 pixels size).

will increase the robustness of the description. In this work, we assume that colour distribution can provide an efficient feature for region description as it is robust to partial occlusion, scaling and region deformation. It is also relatively stable under rotation in depth in certain cases. Besides, colour histograms can be easily quantized into a small number of bins to satisfy the low-computational cost imposed by fast processing. In order to take into account the spatial information, which could be useful to distinguish certain cases, the regions can be masked with a kernel in the spatial domain [4]. Thus, the appearance of the region is described by a set of scalar features, $\{s_i\}_{i=1\dots N}$ which will be obtained from the image region defined by the visual landmark, ζ_v . The value s_n of the n -th bin is defined by

$$s_n = \frac{1}{\mathcal{C}} \sum_{(x,y)_i \in \zeta_v} \mathcal{N}((x,y)_i) \delta(\gamma[I((x,y)_i)] - n) \quad (1)$$

where $\mathcal{N}(\cdot)$ defines a Gaussian-based kernel function which assigns higher weights to the pixels near the centroid than pixels at the borders of the landmark, and \mathcal{C} is a normalization constant ($\mathcal{C} = \sum \mathcal{N}((x,y)_i)$) [4]. δ is the Kronecker delta function and γ is a quantization function, which associates with each observed pixel value $I((x,y)_i)$ a particular bin index. Finally, it must be commented that in our implementation, the CIELab colour space has been chosen to the hierarchical grouping algorithm and also to characterize the colour of the landmark. We have also chosen to quantize the histogram in 16 bins, resulting in a landmark descriptor of 16 x 16 x 16 scalar values.

In order to compute the similarity between kernels (regions), Comaniciu et al [4] propose a metric derived from the Bhattacharyya coefficient. The distance between the discrete distributions \mathbf{p} and \mathbf{q} associated to two visual landmarks is defined as:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{1 - \rho[\hat{\mathbf{p}}, \hat{\mathbf{q}}]} \quad (2)$$

where

$$\rho[\hat{\mathbf{p}}, \hat{\mathbf{q}}] = \sum_{i=1}^m \sqrt{\hat{p}_i \cdot \hat{q}_i} \quad (3)$$

being \hat{p}_i and \hat{q}_i the corresponding bins of the discrete representations \mathbf{p} and \mathbf{q} , respectively.

- *Landmark positioning.* In a mobile robotic navigation framework, each detected landmark must be located in the 3D world. To do that, we can choose a set of stable region points which are subsequently used for referencing the landmark position. These points can be characterized by their Cartesian coordinates provided by a stereoscopic vision system. In order to include these regions as landmarks in a mobile robotic navigation framework, it is usually necessary to estimate the landmark position uncertainties.

Given the two matched pixels associated to a landmark point, the corresponding 3D point coordinates can be computed according to

$$z = \frac{b \cdot f}{d} \quad x = \frac{(u - C_x)}{D_u \cdot f \cdot s_u} z \quad y = \frac{(v - C_y)}{D_v \cdot f} z \quad (4)$$

(C_x, C_y) , b and f being the image center, stereo camera baseline and camera focal length, respectively. $D_u \cdot s_u$ and D_v are the number of pixels per mm for the x and y axis. All of them are calibration parameters. Finally, d and (u, v) are the disparity value and the image coordinates associated to the point. Using a first order approximation, we have

$$\sigma_z^2 \approx \frac{(b \cdot f)^2}{d^4} \sigma_d^2 \rightarrow \sigma_z \approx \frac{\sigma_d}{b \cdot f} z^2 \quad (5)$$

Then, the covariance matrix of this point is

$$\begin{bmatrix} \frac{(u-C_x)^2}{D_u \cdot f \cdot s_u} & \frac{(u-C_x)(v-C_y)}{D_u \cdot f \cdot s_u D_v \cdot f} & \frac{(u-C_x)}{D_u \cdot f \cdot s_u} \\ \frac{(u-C_x)(v-C_y)}{D_u \cdot f \cdot s_u D_v \cdot f} & \frac{(v-C_y)^2}{D_v \cdot f} & \frac{(v-C_y)}{D_v \cdot f} \\ \frac{(u-C_x)}{D_u \cdot f \cdot s_u} & \frac{(v-C_y)}{D_v \cdot f} & 1 \end{bmatrix} \left(\frac{\sigma_d}{b \cdot f} z^2 \right)^2 \quad (6)$$

It can be noted that all values of the covariance matrix depend on the estimate of σ_d . If the point positions are observed with a standard deviation of one pixel, then σ_d can be fixed to $\sqrt{2.0}$ pixels [7].

In this work, we represent landmarks by sets of dominant points extracted from their outer boundaries. If the visual landmark is static and these dominant points are reliably computed, then they will be always located at the same world position. That is, the algorithm could detect or not these points, but if they are detected, their 3D world position will be always the same. This is not true for all characteristic points which can be employed to reference a visual landmark, which could depend on the fact that the landmark will be not occluded (e.g., the center of mass or centroid of the landmark). Next section describes the algorithm we employ to detect these dominant points.

3 Adaptive Approach for Dominant Point Detection

Dominant points are representative features for the object contours which can be mainly detected using polygonal approximation approaches or corner detectors. The goal of the first group of approaches is to represent the contour with the least number of straight–line segments. However, in most cases, the number and position of the detected segments differ with variation of size and orientation of the shape. This limits its use to applications such as partial shape matching or recognition. Corner detectors works by locating the dominant points of the approximating polygon directly through detecting the points with local maxima curvature. Then, they usually conduct two consecutive stages: curvature estimation and local maximum curvatures locating. Fig. 2 presents an example where dominant points are detected by thresholding the estimated curvature function. Figs. 2c-k show the dominant points detected over shapes of the same object which have been affected by affine transformations. Obtained results demonstrate the high stability of the detected dominant points. For this reason, we have adopted a corner detector for dominant point selection.

When a corner detector is employed, dominants points are identified as the points with local extreme curvature. In the continuous case, the curvature of a point is defined as the rate of change between the tangent angle and the arc length. Let $\mathcal{C}(t) = (x(t), y(t))$ be a parametric plane curve. Its curvature function $\kappa(t)$ can be calculated as [17]

$$\kappa(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}} \quad (7)$$

This equation implies that estimating the curvature involves the first and second order directional derivatives of the plane curve coordinates, (\dot{x}, \dot{y}) and (\ddot{x}, \ddot{y}) respectively. This is a problem in the case of computational analysis where the plane curve is represented in a digital form and directional derivatives cannot be exactly computed. To solve this problem, the curvature of each contour point is calculated using the information of the neighboring points. Those neighboring points are designated as the region–of–support of each contour point.

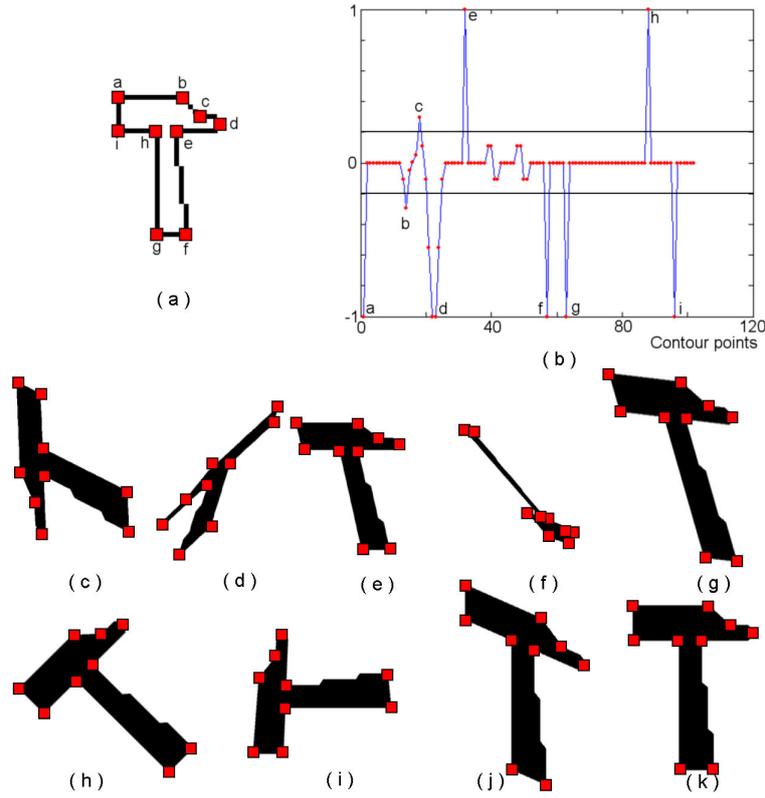


Figure 2: a) Shape #1; b) adaptive curvature function associated to a) and c-k) dominant points detected over transformed shapes. Curvature functions have been estimated using the Bandera et al's algorithm [3] to compute the region-of-support and the cosine measurement for curvature calculation.

From the pioneering paper of Teh and Chin [19], many researchers have argued that the estimation of the shape curvature relies primarily on the precise calculation of the region-of-support associated to each contour point. Let \mathcal{C} be a digital curve defined as a set consisting of N consecutive points

$$\mathcal{C} = \{\mathbf{p}_i = (x_i, y_i)\}_{i=1}^N \quad (8)$$

An adaptive curvature-based approach for dominant point detection must solve three main problems:

- Estimation of the region-of-support associated to each contour point. If both region arms have the same length, then it will be defined by $t[i]$, which will be the same for both region arms. Otherwise, it will be defined by the length of the left and right arms, $t_b[i]$ and $t_f[i]$, respectively.
- Given a region-of-support, computing the curvature value associated to each contour point
- Choosing the set of dominant points from the whole curvature function

Next sections describe the algorithms employed to estimate the region-of-support and curvature associated to each contour point. Once the curvature has been computed for every contour point, dominant points will be obtained by thresholding the curvature values [3].

3.1 Estimation of the Region-of-support

In this work, we use the Bandera et al's proposal to estimate the region-of-support [3]. This approach calculates the maximum length of contour presenting no discontinuities on the right (*forward*) and left (*backward*) sides of the working contour point i : $t_f[i]$ and $t_b[i]$, respectively. $t_f[i]$ is calculated by comparing the Euclidean distance from point i to its $t_f[i]$ -th neighbour ($d(i, i + t_f[i])$) to the length of the contour between both points ($l(i, i + t_f[i])$) which is defined as

$$l(i, i + t_f[i]) = \sum_{j=i}^{t_f[i]-1} d(j, j + 1) \quad (9)$$

Both distances tend to be equal in absence of corners, even if contours are noisy. Otherwise, the Euclidean distance is quite shorter than the contour length. Thus, $t_f[i]$ is the largest value that satisfies

$$l(i, i + t_f[i]) - d(i, i + t_f[i]) < U_k \quad (10)$$

being U_k a user-specified, constant value that depends on the noise level tolerated by the detector. $t_b[i]$ is also set according to Eq. (10), but using $i - t_b[i]$ instead of $i + t_f[i]$. The correct selection of the U_k value is very important. Thus, if the value of U_k is large, $t_f[i]$ and $t_b[i]$ tend to be large and some corners may be missed and, if it is small, $t_f[i]$ and $t_b[i]$ are always very small and the resulting function is noisy. In our tests, this value has been fixed to 0.4.

3.2 Curvature Estimation

Many researchers have used the area of the triangle, formed by the outer boundary points, as the basis for shape representations [1]. The proposed shape recognition system employs a curvature estimator to characterize the shape contour which is based on this triangle-area representation (TAR). Given a shape and, once our proposal have determined the local region-of-support associated to every point of its contour, the process to extract the associated TAR consists of the following steps:

1. Calculation of the local vectors \vec{f}_i and \vec{b}_i associated to each point i . These vectors present the variation in the x and y axis between points i and $i + t_f[i]$, and between i and $i - t_b[i]$. If (x_i, y_i) are the Cartesian coordinates of the point i , the local vectors associated to i are defined as

$$\begin{aligned} \vec{f}_i &= (x_{i+t_f[i]} - x_i, y_{i+t_f[i]} - y_i) = (f_{x_i}, f_{y_i}) \\ \vec{b}_i &= (x_{i-t_b[i]} - x_i, y_{i-t_b[i]} - y_i) = (b_{x_i}, b_{y_i}) \end{aligned} \quad (11)$$

2. Calculation of the TAR associated to each contour point. The signed area of the triangle at contour point i is given by [1]:

$$\kappa_i = \frac{1}{2} \begin{vmatrix} b_{x_i} & b_{y_i} & 1 \\ 0 & 0 & 1 \\ f_{x_i} & f_{y_i} & 1 \end{vmatrix} \quad (12)$$

3. TAR Normalization. The TAR of the whole contour, $\{\kappa_i\}_{i=1}^N$, is normalized by dividing it by its absolute maximum value.

When the contour is traversed in counter clockwise direction, positive, negative and zero values of TAR mean convex, concave and straight-line points, respectively.

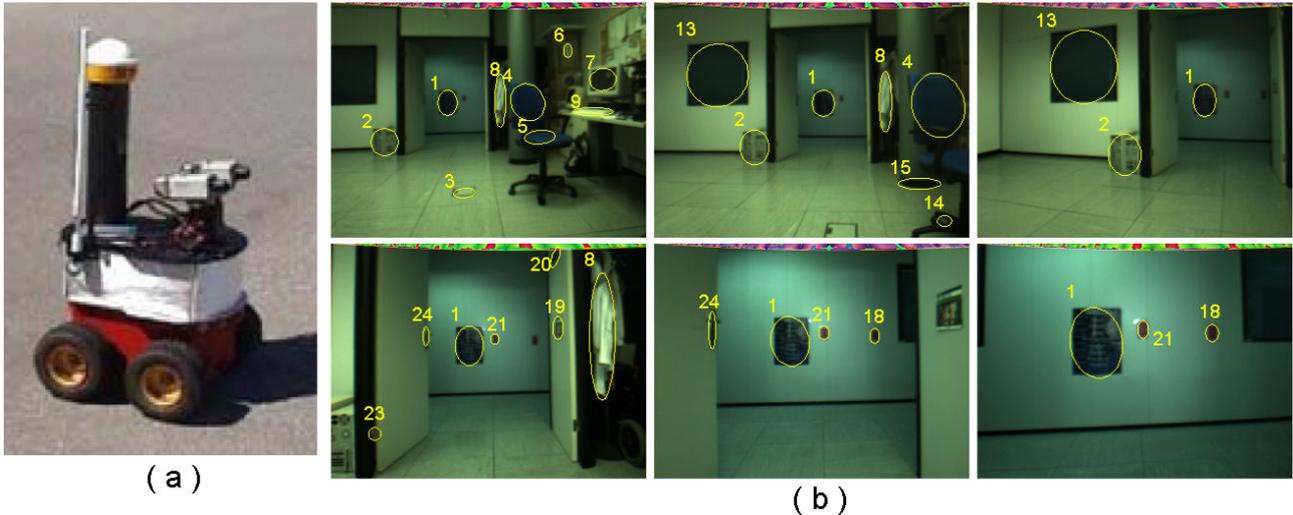


Figure 3: a) REX: a Pioneer 2AT robot from ActivMedia equipped with a stereo vision system, GPS and laser; and b) some acquired left images (detected landmarks are represented by the ellipses drawn over the images).

4 Experimental Results

To test the validity of the proposed system, data was collected with an ActiveMedia Pioneer 2AT robot mounted with an stereoscopic camera. The robot was driven through different environments while capturing real-life stereo images. The stereo head is the STH-MDCS from Videre Design: a compact, low-power colour digital stereo head with an IEEE 1394 digital interface. It consists of two 1.3 megapixel, progressive scan CMOS imagers mounted in a rigid body, and a 1394 peripheral interface module, joined in an integral unit. The camera was mounted at the front and top of the vehicle at a constant orientation, looking forward. Images obtained were restricted to 640x480 pixels. Figs. 3a-b show the robot used in the experiments and several acquired images. Detected landmarks have been represented by the ellipses drawn over the images. It can be noted that we have also overimposed a landmark index. This index has been obtained by matching landmark kernel-based descriptors between consecutive frames.

4.1 Visual Landmark Detection

To quantitatively check the viewpoint invariance of the detector, we compare our method to other similar approaches using the protocol proposed by Mikolajczyk et al [16]. The comparison database¹ is composed by eight different image sets that represent five changes in imaging conditions (viewpoint changes, scaling, image blur, jpeg compression and illumination changes). A ground truth homography transformations are provided between first images of the sequence (reference image) and the other images. Fig. 4 shows an example from each image set. It must be noted that the set of parameters employed by the proposed approach has not been modified to deal with the different image sets.

To evaluate the detection ability, the repeatability score is employed [16]. The objective of this test is to measure how many of the detected regions are found in images under different transformations, relative to the lowest total number of regions detected (where only the part of the image that is visible

¹<http://www.robots.ox.ac.uk/vgg/research/affine/>

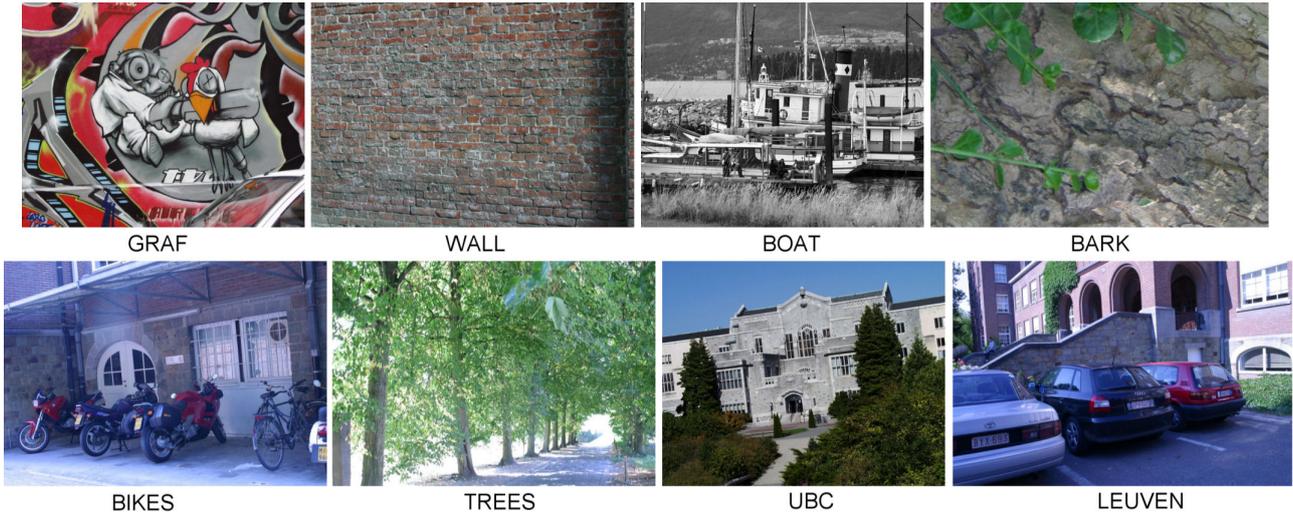


Figure 4: Image examples of the eight sets used for comparison purposes.

in both images is taken into account). The measure of repeatability is the relative amount of overlap between regions detected in the reference image and in the other image. This region is projected onto the reference image using the homography relating the images. It must be noted that the output for our detector is a set of arbitrarily shaped regions. However, for the purpose of the comparisons using this protocol, the output region of all detectors are represented by an ellipse. In our case, ellipses which have the same first and second moments as the detected regions are used to approximate them. The proposed detector is compared to the Hessian–Affine detector [14], the *maximally stable extremal region* detector (MSER) [12] and the *intensity extrema-based region* detector (IBR) [20]. These approaches have been selected because they obtain the highest scores in many cases in the work of [16]. Besides, as it is mentioned in this work, the MSER and IBR detectors are the best choices if only a very small number of matches is needed (e.g. to localize a mobile robot). For all experiments, the default parameters given by the authors are used for each detector. The repeatability for several sets of images are illustrated in Fig. 5. Results for the rest of sets looks similar. They show that the proposed detector ranks similar to the MSER. In the different sequences, the proposed approach only detects a reduced set of regions and the thresholds can be set very sharply, resulting in very stable regions. Besides, the hierarchical processing allows the method to detect these regions at a very low computational time (e.g., it takes less than 500 ms in a Pentium 4.2 GHz Linux PC, for the image shown in Fig. 1a, which is 800 x 640 pixels).

4.2 Kernel-based Representation

The kernel-based descriptor is evaluated using the recall-precision criterion for image pairs, i.e. the number of correct and false matches between two images [15]. Recall is defined as the number of correctly matched regions with respect to the number of corresponding regions between two images of the same scene. The precision is defined as the number of correct matches with respect to the total number of matches. The results are represented with recall versus 1-precision. Fig. 6 shows the results for several sets of images. Landmarkss have been detected using the perceptual–grouping approach described in Section 2. Two visual landmarks are matched if the distance between their descriptors is below a threshold U . The value of this threshold is varied to obtain the curves (see [15] for further details). In Fig. 6, the kernel-based descriptor is compared with the SIFT [8] and the

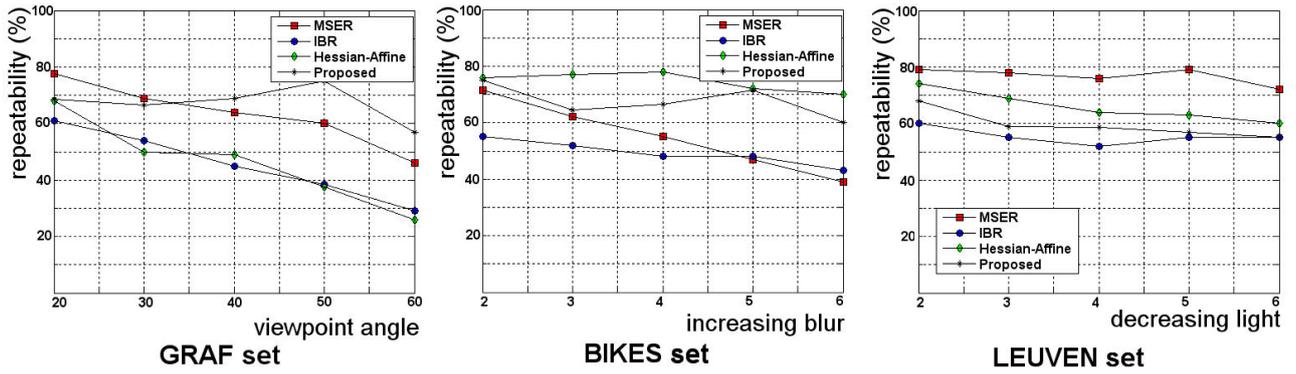


Figure 5: Repeatability scores for GRAF, BIKES and LEUVEN sequences (see Fig. 4).

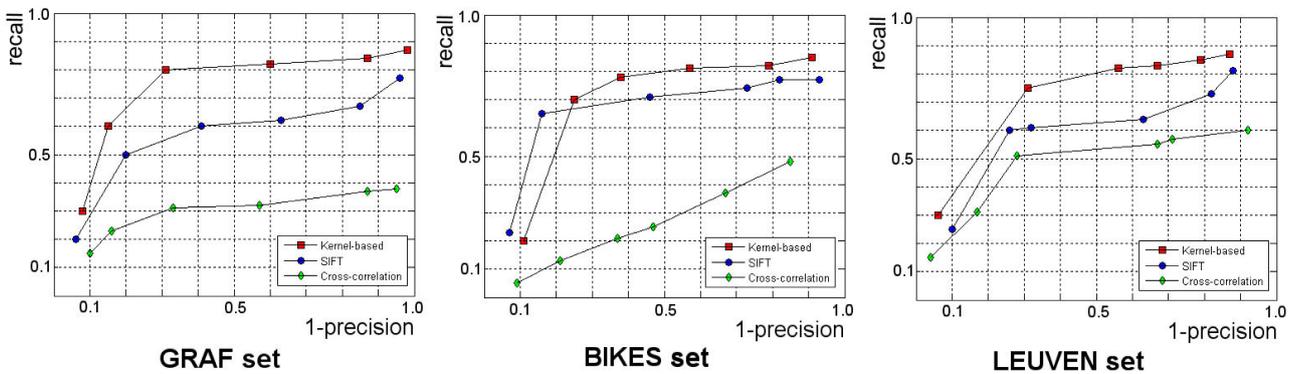


Figure 6: Recall vs. 1-precision curves for GRAF, BIKES and LEUVEN sequences (see Fig. 4).

cross-correlation (evaluated for a path of 11×11 pixels centered at the center of mass of the detected landmark). From the results, it can be noted that the kernel-based descriptor performs better than the rest of descriptors. The number of landmarks is significantly low, and they are not overlapped. Besides, although the textured scenes contain similar motifs, the regions capture distinctive image variations. For these reasons, distribution-based descriptors like the kernel-based one or the SIFT, exhibit a good performance. On the other hand, the size of the kernel-based descriptor is significantly larger than the rest of descriptors. This implies more computational time and storage resources, which are compensated by its better performance, specially when dealing with real acquired images.

4.3 Corner-based Description

To evaluate the robustness to affine transformations of the dominant point detector, a similarity measure between affine-transformed landmark shapes and the original, have been computed. In our tests, this measure is provided by the correlation between the curvature functions associated to both shapes. For comparison purposes, other approaches have been also evaluated. Table 2 shows the obtained results for some of the shapes on Fig. 2. Evaluated approaches have been classified according to the algorithm they employ to estimate the region-of-support and to compute the curvature (see [2] for further details).

From all conducted tests, it has been concluded that the best adaptive approach for corner detection

Region-of-support	Curvature	Fig. 2c	Fig. 2d	Fig. 2e	Fig. 2f
Bandera et al	Cosine	85.2	71.3	68.0	57.5
Marji and Siy	Cosine	67.4	64.5	60.4	52.6
Bandera et al	TAR	99.9	71.9	70.3	99.9
Marji and Siy	TAR	64.6	77.1	77.7	93.6
Teh and Chin	Cosine	96.1	87.1	95.2	76.5
Ray and Ray	Cosine	94.2	65.7	59.0	47.9
Wu	k -cosine	89.1	73.4	69.9	62.5

Table 2: Comparison of adaptive approaches.

combines the Bandera et al’s proposal for estimating the region-of-support [3] and the TAR approach for curvature computing.

5 Conclusions and Future Work

Landmark-based maps constitute one of the most popular choices for environment mapping in mobile robotics. In this paper, we describe a visual landmark detection and description system. The detection stage is achieved using a perceptual-based hierarchical algorithm. On the other hand, dominant points detected over the landmark contour are used to locate the entire landmark in the 3D real world. The visual appearance of detected landmarks is characterized by a kernel-based representation.

Among other issues, the success of a landmark-based representation is conditioned on the availability of fast and reliable algorithms which will be able to provide viewpoint-independent detection and recognition of landmarks. If we assume the local planarity of landmarks and adequacy of the affine approximation of the geometric changes induced by the movement of the robot’s vision system, any photometrically normalized image measurement expressed in a local affine frame coordinate system is viewpoint-invariant. Characterizing the landmark by a set of dominant points allows to define a local region-centered, affine-invariant coordinate system, which will permit to solve all six degrees of freedom which an affine transformation possesses. It must be noted that affine coordinate systems cannot be constructed directly from interest point detectors, like the SIFT [8]. In this sense, our proposal is more related to region-based detectors (e.g. the MSER [12] or the IBR [20] detectors). However, contrary to these proposals, the proposed approach works in the scale-space provided by the irregular pyramid. In any case, extensive tests must be conducted to evaluate the advantages of the proposal with respect to other region-based detectors. Future works will be also focused on the development of an algorithm which allows to track the landmarks in presence of occlusion and false detections, and on the integration of this system into a simultaneous localization and mapping (SLAM) framework.

References

- [1] N. Alajlan, I. El Rube, M. Kamel, and G. Freeman. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recognition*, 40:1911–1920, 2007.
- [2] E. Antúnez, A. Bandera, and R. Marfil. Comparison of adaptive approaches for dominant points detection. *Computer Vision Winter Workshop 2009*, 2009.

- [3] A. Bandera, C. Urdiales, F. Arrebola, and F. Sandoval. Corner detection by means of adaptively estimated curvature function. *Electronics Letters*, 36(2):124–126, 2000.
- [4] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(5):564–577, 2003.
- [5] Y. Haxhimusa, R. Glantz, M. Saib, G. Langs, and W.G. Kropatsch. Logarithmic tapering graph pyramid. *DAGM2002 (Lecture Notes of Computer Science)*, 2449:117–124, 2002.
- [6] J.M. Jolion. Stochastic pyramid revisited. *Pattern Recognition Letters*, 24(8):1035–1042, 2003.
- [7] T. Lemaire, C. Berger, I. Jung, and S. Lacroix. Vision-based slam: stereo and monocular approaches. *Int. Journal of Computer Vision*, 74(3):343–364, 2007.
- [8] D.G. Lowe. Object recognition from local scale invariant features. *Proc. 7th Int. Conf. on Computer Vision*, pages 1150–1157, 1999.
- [9] R. Madhavan and H.F. Durrant-Whyte. Natural landmark-based autonomous vehicle navigation. *Robotics and Autonomous Systems*, 46:79–95, 2004.
- [10] R. Marfil, A. Bandera, and F. Sandoval. Perception-based image segmentation using the bounded irregular pyramid. *DAGM2007 (Lecture Notes of Computer Science)*, 4713:244–253, 2007.
- [11] R. Marfil, L. Molina-Tanco, A. Bandera, and F. Sandoval. The construction of bounded irregular pyramids with a union-find decimation process. *GbrPR2007 (Lecture Notes of Computer Science)*, 4538:307–318, 2007.
- [12] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Proc. British Machine Vision Conference*, pages 384–393, 2002.
- [13] P. Meer. Stochastic image pyramids. *Computer Vision, Graphics and Image Processing*, 45:269–294, 1989.
- [14] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. *Proc. 7th European Conference on Computer Vision*, pages 128–142, 2002.
- [15] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(10):1615–1630, 2005.
- [16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. Journal Computer Vision*, 65:43–72, 2006.
- [17] F. Mokhtarian and A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Analysis and Machine Intell.*, 8(1):34–43, 1986.
- [18] C. Siagian and L. Itti. Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 29(2):300–312, 2007.
- [19] C.H. Teh and R.T. Chin. On the detection of dominant points on digital curves. *IEEE Trans. on Pattern Anal. Machine Intell.*, 11:859–872, 1989.
- [20] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *Int. Journal on Computer Vision*, 59(1):61–85, 2004.