

Adaptive Dominant Points Detector for Visual Landmarks Description*

Esther Antúnez¹, Rebeca Marfil², Antonio Bandera², and Walter G. Kropatsch¹

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

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

Abstract

This paper describes a visual landmark detection and description system for mobile robotic environment mapping. At the detection stage, the proposal employs a perceptual-based hierarchical mechanism which has been previously used to successfully solve the 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 considered and evaluated in this paper, it is specifically focused on the dominant points detection. Hence, this paper evaluates the performance of a fast curvature-based approach for adaptive detection and its robustness to shape affine transformations.

1. Introduction

Reliable navigation is an essential component of an autonomous mobile robot which typically implies to represent 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, 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 [12]. 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 landmarks detection and description system for mobile robotic 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). This pyramid and its application to low-level image segmentation have been published in conference papers [5, ?]. However, the BIP presents an inherent problem which can damage its application on this framework:

*This work has been partially granted by the Spanish Junta de Andalucía project P07-TIC-03106.

it provides an image segmentation which varies when this image is shifted slightly [5]. 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, a set of image regions which satisfies certain criteria are 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 module consists of three stages. Firstly, a hierarchical perceptual grouping algorithm is applied to perform a domain-independent segmentation of the image pixels into regions which will be mainly based on properties like the proximity, similarity, closure or continuity. 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 (*visual landmarks validation stage*). 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. 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). Finally, obtained landmarks are characterized by the 3D position of a set of dominant points detected over its outer contour. A kernel-based representation [3] is adopted to describe its internal colour distribution.

3. Adaptive Approach for Shape Description

Among the elementary visual features which can be used for to describe the landmark, shape is the only feature which is usually able to reveal its identity [4]. Many shape descriptors have been proposed over the past decades. Among them, contours constitute one of the most employed descriptors due to their semantically rich nature [6]. In this work, we will focus on those approaches that attempts to represent shapes with a limited number of dominant points located along their contours. Dominant points are considered as representative features for the object contours and they can be mainly detected using polygonal approximation approaches and 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. ?? presents an example where dominant points are detected by thresholding the estimated curvature function. Fig.

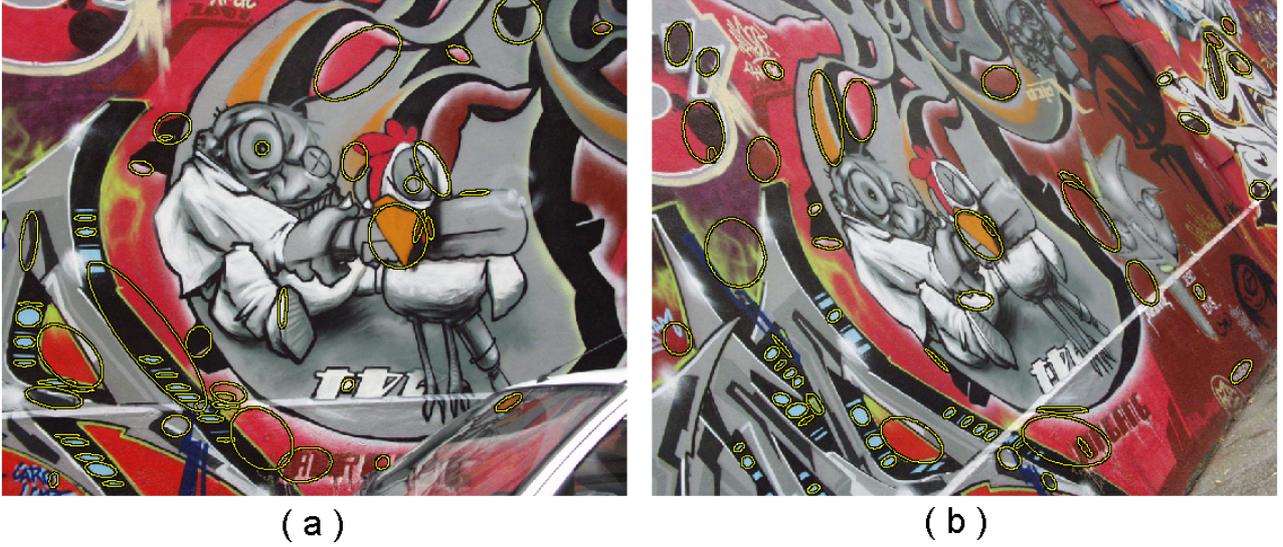


Figure 1. Regions generated by the proposed detector on two images taken from different viewpoints. The ellipses show the original detection size.

Figure 1c shows that linking detected dominant points by line segments it is possible to obtain a polygonal approximation which can be easily related to the original shape. In this paper, we have adopted a corner detector for dominant points selection.

When a corner detector is employed, dominant 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 [11]

$$\kappa(t) = \frac{\dot{x}(t)\ddot{y}(t) - \ddot{x}(t)\dot{y}(t)}{(\dot{x}(t)^2 + \dot{y}(t)^2)^{3/2}} \quad (1)$$

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.

From the pioneering paper of Teh and Chin [13], 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

$$\mathcal{C} = \{\mathbf{p}_i = (x_i, y_i)\}_{i=1}^N \quad (2)$$

An adaptive curvature-based approach for dominant points 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 k_i , which will be the same for both region arms. Otherwise, it will be defined by the length of the left and right arms, k_i^b and k_i^f , 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 [2].

3.1. Estimation of the region–of–support

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

$$\begin{aligned}\vec{f}_i &= (x_{i+t_f[i]} - x_i, y_{i+t_f[i]} - y_i) = (f_{x_i}, f_{y_i}) \\ \vec{b}_i &= (x_{i-t_b[i]} - x_i, y_{i-t_b[i]} - y_i) = (b_{x_i}, b_{y_i})\end{aligned}\quad (3)$$

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

$$\kappa_i = \frac{1}{2} \begin{vmatrix} b_{x_i} & b_{y_i} & 1 \\ 0 & 0 & 1 \\ f_{x_i} & f_{y_i} & 1 \end{vmatrix}\quad (4)$$

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.

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.

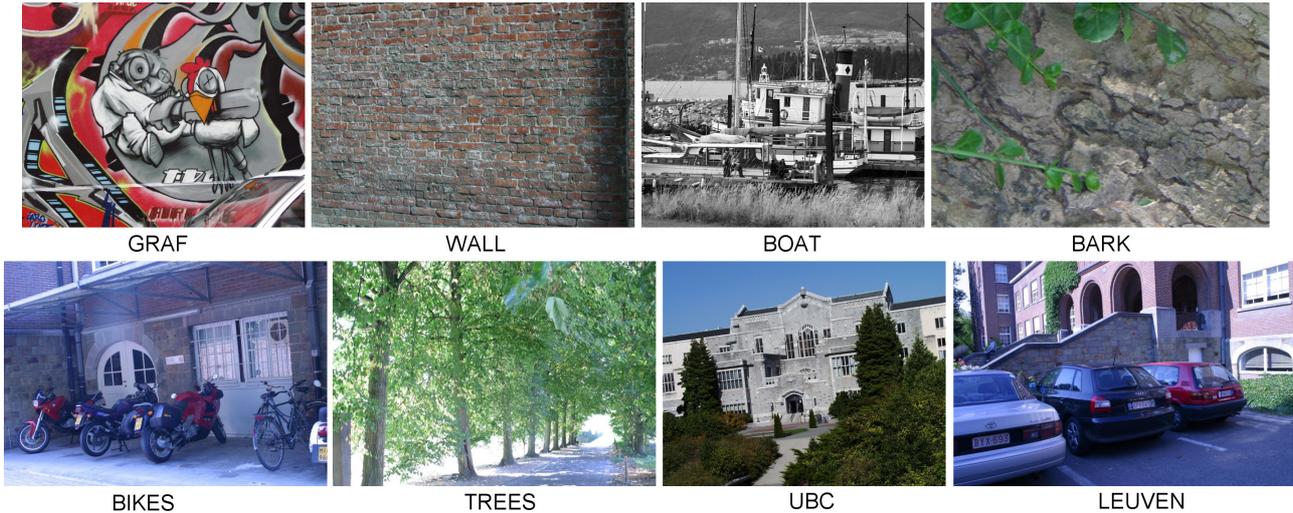


Figure 2. Image examples of the eight sets used for comparison purposes.

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 [10]. 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. 2 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 [10]. 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 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 [8], the *maximally stable extremal region* detector (MSER) [7] and the *intensity extrema-based region* detector (IBR) [14]. These approaches have been selected because they obtain the highest scores in many cases in the work of [10]. 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. for 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. 3. 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

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

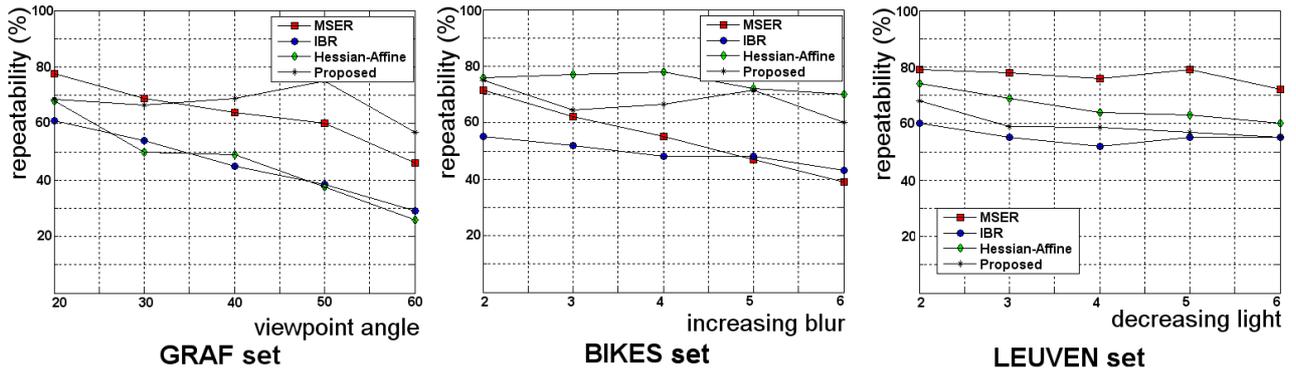


Figure 3. Repeatability scores for GRAF, BIKES and LEUVEN sequences (see Fig. 2).

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 [9]. 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. 4 shows the results for several sets of images. Regions have been detected using the perceptual-grouping approach described in Section 2. Two regions 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 [9] for further details). In Fig. 4, the kernel-based descriptor is compared with the SIFT [?] and the cross-correlation (evaluated for a path of 11 x 11 pixels centered at the center of mass of the detected region). From the results, it can be noted that the kernel-based descriptor performs better than the rest of descriptors. The number of regions is significantly low, and this implies that regions are usually 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

5. Conclusions and Future Work

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.

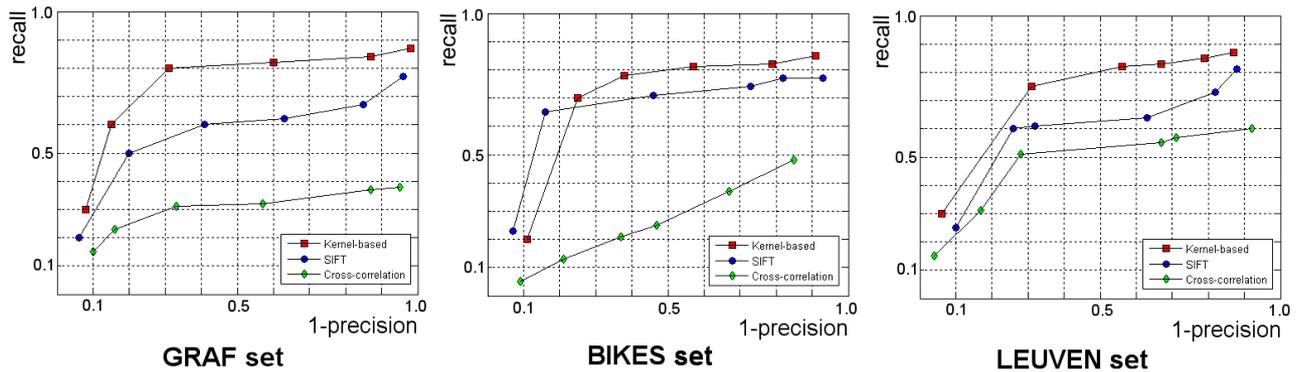


Figure 4. Recall vs. 1-precision curves for GRAF, BIKES and LEUVEN sequences (see Fig. 2).

- [2] 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.
- [3] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(5):564–577, 2003.
- [4] S. Lambert, E. de Leau, and L. Vuurpijl. Using pen-based outlines for object-based annotation and image-based queries. In *Proc. 3rd Int. Conf. Visual Information and Information Systems*, pages 585–592, 1999.
- [5] D.G. Lowe. Object recognition from local scale invariant features. *Proc. 7th Int. Conf. on Computer Vision*, pages 1150–1157, 1999.
- [6] R. Marfil, L. Molina-Tanco, A. Bandera, and F. Sandoval. The construction of bounded irregular pyramids with a union-find decimation process. *Lecture Notes of Computer Vision*, 4538:307–318, 2007.
- [7] M. Marji and P. Siy. A new algorithm for dominant points detection and polygonization of digital curves. *Pattern Recognition*, 36:2239–2251, 2003.
- [8] 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.
- [9] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. *Proc. 7th European Conference on Computer Vision*, pages 128–142, 2002.
- [10] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(10):1615–1630, 2005.
- [11] 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.
- [12] 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.

- [13] 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.
- [14] 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.
- [15] 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.