Detection of Natural Landmarks through Multiscale Opponent Features

Eduardo Todt[1], Carme Torras

Institut de Robòtica i Informàtica Industrial (CSIC-UPC)
Gran Capità, 2-4 (edifici Nexus), 08034-Barcelona, Spain
todt@ieee.org, torras@iri.upc.es

Abstract

This work presents a landmark detection system for a walking robot, which has to operate in unknown unstructured outdoor environments. Most landmark detection approaches are not adequate for this application, since they rely on either structured information or a priori knowledge about the landmarks. Instead, the proposed system makes use of visual saliency concepts stemming from studies of animal and human perception. Thus, biologically-inspired opponent features (in color and orientation) are searched for at different resolution levels. The implementation, however, does not try to mimic nature, but rather to be as computationally efficient as possible. Thus, salient image regions ranging from relatively small to big sizes are detected using multiscale comparison techniques, based on pyramidal filtering. The experimental results obtained show that visual saliency permits detecting reliable natural landmarks without a priori knowledge about their characteristics or location.

1. Introduction

This work deals with the problem of detecting natural visual landmarks in outdoor environments. Natural landmarks are objects or patterns already present in the environment that can be used as reference marks to identify places in the world, like trees in a field, rocks, and buildings. In opposition to them, artificial landmarks are those inserted deliberately in the environment to serve as reference marks, such as bar code signs, radio bacons, and geometrical figures.

When dealing with visual landmarks in outdoor environments, color and texture are two of the most important features, since they are adequate to characterize unstructured forms [21]. The main difficulty in their utilization as recognition features is that the visual information acquired is strongly dependent on the intensity, direction, and spectral power distribution of the illumination, which change with sun position and atmospheric condition [22], [9].

In order to increase robustness to these adverse conditions, the original Red, Green, and Blue (*RGB*) color components are often transformed to another color space, like the normalized color space (chromaticity coordinates), which is decoupled from the luminance, or the Hue, Intensity and Saturation (*HIS*), well-suited to human perception representation [22], [17].

Additionally to the illumination dependency of color perception, the robot position and orientation at the observation time influence the visibility, distance, and orientation of the landmarks relative to the robot. Since for mobile robots these conditions are not fixed along time, it is desirable to deal with features as independent as possible from the robot's movements, in order to achieve robust landmark recognition.

Most published works on visual landmark-based navigation are related to indoor environments (e.g., [7], [14], [8], [20], [6]), which have controlled illumination and structured or artificial landmarks available.

In outdoor environments, the existence of a priori maps with the landmark locations is frequently assumed, as well as the availability of landmark models. Nasr and Bhanu [15] use model-based vision, assuming explicit knowledge about the maps, 3-D models of landmarks and a camera model. They adopt several features for the landmark descriptions, some range-dependent, such as size, length, and width, and others range-independent, like color, perimeter squared over area, length over width, and shape. Talluri and Aggarval [19] use the horizon line contour as landmark, matching the detected contours against a known map. Cozman [3] also works with the horizon line, but instead of recognizing contours he uses mountain peaks as landmarks. Evans, Smith and Lucas [4] use a priori models of the landmarks, which are identified by their color, texture and shape features. Takeuchi and Herbert [18] use normalized red color distribution, vertical and

---

[1] Professor at Pontifícia Universidade Católica do Rio Grande do Sul, Brazil
Currently PhD student at Universitat Politècnica de Catalunya

976

horizontal edges, and texture of images in urban areas to characterize buildings, that are taken as landmarks. Zheng, Barth and Tsuji [23] detect landmarks along the borders of a route, based on the contrasts of the color and intensity features.

It can be observed that the reported landmark characterization methods for outdoor environments usually rely on a priori models. Instead of following this approach, this work proposes the usage of visual saliency techniques to select good landmark candidates in the images, without a priori knowledge about the landmarks. The presented visual saliency is based on multiscale opponent features, obtained from intensity, color and texture characteristics.

2. Visual saliency

In the case where there is no a priori knowledge of what things in the scenes will be used as landmarks, some criterion is needed to decide what regions in the images could be considered as representing potential landmarks.

Recent theories in active and purposive vision indicate that potential landmark locations could be detected using a visual attention mechanism. It has also been observed that when human perceivers are trying to build or recover the description of a scene, their attention is focused on certain relevant regions in the scene [10], [11].

One important concept in visual attention is the saliency map. This map represents the regions of the image corresponding to distinctive areas with large values, whereas the non-distinctive regions are indicated by small values. Usually a saliency map is constructed from a linear combination of a set of features, which is application dependent. In addition, the use of a visual attention mechanism bypasses the task of image segmentation for object delineation, which does not always produce accurate results [20].

Two interesting works on visual attention that intend to model and predict human visual search performance in color images are presented by Itti, Koch, and Niebur [12] and Fernandez-Vidal et al. [5]. Their systems build topographical saliency maps from a combination of multiscale image features; specifically, intensity, opponent colors and orientation. Contrarily to these works, the landmark detection system here described does not attempt to emulate human visual search mechanisms, but, inspired by their approach, uses the same concept of saliency maps to locate good landmark candidates.

3. The landmark detection system

Each acquired color image has resolution of 512x384 pixels and is represented in conventional camera RGB space, with 24 bits per pixel (8 bits for each channel

RGB). The landmark detection system computes the visual saliencies based on the opponency of intensities, colors and main orientations present in the images.

Inspired on biological models of vision [1], the detection of the visual saliencies is made through the computation of center-surround differences, i.e., comparing regions of the visual field with their respective neighborhoods. For the intensity feature the bright-dark opponency is considered, meaning that dark centers with bright surrounds or bright centers with dark surrounds are salient. For the color features, red centers with green surrounds, or green centers with red surrounds are considered salient; the same concept is valid for the blue-yellow color pair. Finally, the local orientation features are computed using Gabor filters, oriented at four different angles, 0°, 45°, 90° and 135°. For all considered features, the greater is the contrast between the center and surround regions, the greater is the correspondent saliency.

Considering that in outdoor environments the intensity of the illumination can change substantially along time, altering the color perception of the objects in the RGB space, it is convenient to decouple the color information from the illumination intensity. This is achieved converting the image representation from the RGB space to rgb chromaticity space [22].

To compute the desired color opponencies it is necessary to convert the rgb space to another space which gets full response for the saturated colors red, green, blue, and yellow (R´, G´, B´, Y´). Following the proposal presented in [12], this color space is calculated by

The orientation features are computed on the monochromatic intensity image obtained at the beginning

$$R' = r - (g + b) / 2$$
$$G' = g - (r + b) / 2$$
$$B' = b - (r + g) / 2$$
$$Y' = (r + g) / 2 - |r - g| - b$$

of the processing of each image. For each orientation, an efficient Gabor filtered approximation is computed using a convolution with 11-tap filter masks [16].

Each feature (intensity, opponent colors, and orientations) is represented in nine spatial scales, through the use of Gaussian pyramids [2]. In these pyramids, each level is obtained by a low-pass filtering operation on the preceding level, followed by a subsampling of factor two in each dimension. All pixels at any level have equal contribution to the subsequent level. The center-surround differences are computed by the comparison between fine and coarse scales of the pyramids, as presented in [12]. The center corresponds to a feature at a fine scale, and the surround corresponds to a feature at a coarse scale, represented with lower resolution. Figure 1 shows the multiresolution pyramid for the intensity feature, with each level expanded to the size of the first level, in order

to facilitate the comparison procedure. It is possible to observe the center-surround characteristic for pixels corresponding to the same location, but situated in different pyramid levels.
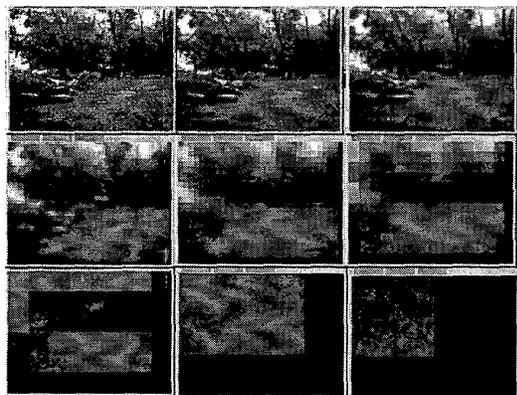

Figure 1. Multiresolution Gaussian pyramid

The low-pass filtering is speeded up taking advantage of the Gaussian filters separability, using a double convolution with a 1-D mask [13]. An important property of the pyramid implementation is that it allows multiscale opponent computations. By comparing centers and surrounds located at high-resolution levels, visual saliencies of relative small targets are found, while operating with lower resolution levels permits finding relatively large salient regions. With this technique, it is possible to detect visual salient objects within a wide size range, e.g., from small trees to big mountains.

The computation of center-surround differences for the intensity, opponent colors and orientation features at each scale combination results in several partial visual saliency maps. These maps are normalized by their local maxima and then combined into a global visual saliency map. The resulting global saliency map has the cues to identify potential landmarks. The image regions corresponding to these potential landmarks are then analyzed with the objective of obtaining visual signatures, capable of identifying them as an existing or a new landmark.

4. Results

All image-processing operations have been implemented using the Matrox® Imaging Library (MIL) version 5.0, in Windows™ 95 environment. The performance tests were evaluated using an Intel® Pentium MMX™, 200MHz PC, with 80Mb of DRAM. Superior microprocessor types were not considered for the tests due to the interest of using this system in a mobile robot, equipped with an easily available off-the-shelf PC/104 computer. The following figures illustrate an example of landmark detection performed by the system. Figure 2

shows an image acquired by the robot, and, in Figure 3, salient regions are indicated. Note that candidate landmarks are found in the most informative areas, namely ground-sky separation and, especially, the rock pattern at the left, which is undoubtedly the most distinctive feature in the image.

Figure 2. Original image acquired by the robot

Figure 3. Resultant landmark candidates

The complete processing of each image takes 18 seconds. The observed bottleneck is the computation of the center-surround differences, which is executed at six different scales for each opponent feature (red-green, blue-yellow, intensity, and four orientations). Although the experimental computation time is compatible with the movement speed of a walking robot, this performance can be improved with some adaptive opponent feature selection, avoiding the overhead of considering features not so relevant in the current environment of the robot.

5. Conclusions

Based on the present results, it seems promising that a multiscale opponent feature mechanism can be used to

978

detect good candidates for landmarks in outdoor environments, avoiding an exhaustive search in the images, in a way independent of predefined models. The main contribution of this work is to apply the visual saliency concepts to natural landmark detection, without an a priori knowledge of the landmarks characteristics.

The main practical problems detected with the experiments are related to the presence of large landscape elements (e.g., rivers, roads), that although may be visually salient, cannot always be used as punctual landmarks because they do not identify a unique location or orientation in space. The idea to solve this problem is to characterize the elements with a measure of their texture uniformity, which can provide some indication about the spatial distribution of each element. Also it is necessary to include in the system a sky elimination procedure, necessary to avoid considering clouds as landmark candidates. Finally, saliencies at the borders of the images are not reliable, due to the lack of data to compute the filter convolutions and surround information.

The selected color space of chromaticity coordinates has proven to be robust in front of illumination intensity changes, as desired. However, this color representation is affected by the spectral power distribution (SPD) of the light source. Since in outdoor environments there is a significant variation in the SPD according to the sun position (direct sunlight has more blue components than sunset light), experiments with alternative color spaces more robust to SPD changes will be considered in future work. Gevers and Smeulders [9] propose an interesting color space with this characteristic, where a logarithmic color ratio is computed between neighboring pixels, resulting in a transformation of the image independent of the illumination intensity and SPD.

Acknowledgments

6. References

[1] Bruce, V., Green, P.R., and Georgeson, M.A., *Visual Perception*, Psychology Press, United Kingdom, 1997.
[2] Burt, P.J., "The pyramid as a structure for efficient computation", in *Multiresolution image process and analysis*, ed. Rosenfeld, A., Springer-Verlag, Heidelberg, 1984, 6-35.
[3] Cozman, F., "Decision Making Based on Convex Sets of Probability Distributions: Quasi-Bayesian Networks and Outdoor Visual Position Estimation", *Tech. Rep. CMU-RI-TR-97-49*, Carnegie Mellon University, Pittsburgh, U.S.A., 1997.
[4] Evans, J.C., Smith, J.S., and Lucas, J., "Improving navigation using active landmark recognition", *IEEE*

Colloquium on Autonomous Underwater Vehicles and their Systems - Recent Developments, 1996, 8/1-8/6.
[5] Fdez-Vidal, X.R., Garcia, J.A., Fdez-Valdivia, J., and Rodriguez-Sánchez, R., "Computing visual target distinctiveness through integral opponent-color features", *VIII National Symposium on Pattern Recognition and Image Analysis*, Vol. I, Bilbao, May 1999, 429-436.
[6] Franz, M.O., Schölkopf, B., Mallot, H.A., and Bülthoff, H.H., "Learning view graphs for robot navigation", *Autonomous Robots* 5, 1998, 111-125.
[7] Fukuda, T., Yokoyama, Y., Abe, Y., and Tanaka, K., "Navigation system based on ceiling landmark recognition for autonomous mobile robot", *IEEE ICRA*, 1996, 1720-1725.
[8] Gaussier, P., Joulain, C., Zrehen, S., Banquet, J.P., and Revel, A., "Visual Navigation in an open environment without map", *IEEE/RSJ IROS*, 1997, 545-550.
[9] Gevers, T. and Smeulders, A.W.M., "Color-based object recognition", *Pattern Recognition*, 32(3), 1999, 453-464.
[10] Grossberg, S., Mingolla, E., Ross, W., "A neural theory on attentive visual search: interactions of boundary, surface, spatial, and object representations", *Psychological Review*, 101(3), 1994, 470-489.
[11] Humphreys, G.W., *Visual cognition: computational, experimental and neuropsychological perspectives*, Lawrence Erlbaum Associates Ltd. Publishers, Exeter, 1989
[12] Itti, L., Koch, C., and 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.
[13] Jain, R., Kasturi, R., and Schunck, B.G., *Machine Vision*, McGraw-Hill, U.S.A., 1995.
[14] Jones, S.D., Andersen, C.S., and Crowley, J.L., "Appearance based processes for visual navigation", *IEEE/RSJ IROS*, 1997, 551-557.
[15] Nasr, H. and Bhanu, B., "Landmark recognition for autonomous mobile robots", *IEEE ICRA*, 1988, pp.1218-1223.
[16] Nestares, O., Navarro, R., Portilla, J., and Tabernero, A., "Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions", *J. Electronic Imaging*, 7, 1998, 166-173.
[17] Sangwine, S.J. and Horne, R.E.N., *The color image processing handbook*, Chapman & Hall, London, 1998.
[18] Takeuchi, Y. and Hebert, M., "Evaluation of Image-Based Landmark Recognition Techniques", *Tech. Report CMU-RI-TR-98-20*, Carnegie Mellon Univ., Pittsburgh, U.S.A., 1998.
[19] Talluri, R., and Aggarwal, J.K., "Position estimation for an Autonomous Mobile Robot in an Outdoor Environment", *IEEE Trans. on Robotics and Autom.*, 8(5), 1992, 573-584.
[20] Trahanias, P.E., Velissaris, S., and Garavelos, T., "Visual Landmark Extraction and Recognition for Autonomous Robot Navigation", *IEEE/RSJ IROS*, 1997, 1036-1042.
[21] Wouwer, G.V., Scheunders, P., Livens, S., and Dyck, D.V., "Wavelet correlation signatures for color texture characterization", *Pattern Recognition*, 32(3), 1999, 443-451.
[22] Wyszecki, G. and Stiles, W.S., *Color Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley and Sons, U.S.A., 1982.
[23] Zheng, J.Y., Barth, M., and Tsuji, S., "Autonomous Landmark Selection for Route Recognition by a Mobile Robot", *IEEE ICRA*, 1991, 2004-2009.