Color-contrast landmark detection and encoding in outdoor images

Eduardo Todt[1] and Carme Torras[2]

[1]Faculty of Informatics, PUCRS, Av. Ipiranga, 6681,
90619-900 Porto Alegre, Brazil
todt@ieee.org
[2]Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6,
08028 Barcelona, Spain
torras@iri.upc.edu

Abstract. This paper describes a system to extract salient regions from an outdoor image and match them against a database of previously acquired landmarks. Region saliency is based mainly on color contrast, although intensity and texture orientation are also taken into account. Remarkably, color constancy is embedded in the saliency detection process through a novel color-ratio algorithm that makes the system robust to illumination changes, so common in outdoor environments. A region is characterized by a combination of its saliency and its color distribution in chromaticity space. The newly acquired landmarks are compared with those already stored in a database, through a quadratic distance metric of their characterizations. Experimentation with a database containing 68 natural landmarks acquired with the system yielded good recognition results, in terms of both recall and rank indices. However, the discrimination between landmarks should be improved to avoid false positives, as suggested by the low precision index.

I. Introduction

The extraction of reliable visual landmarks in outdoor unstructured environments is still an open research problem. Our motivation for working on it comes from robot navigation, but the main issues concern also other fields, such as scene analysis and image indexing and retrieval from databases. Most existing feature extraction approaches are not adequate for this type of environments, since they rely on either structured information from non-deformable objects [3, 8], or a priori knowledge about the landmarks [1].

We have been pursuing a saliency-based approach to spot image regions with potential to represent good landmarks [13, 14], following biologically-inspired works on visual attention [7]. In [14], we introduced a way to embed color constancy within saliency computation, which showed to be faster and more stable than ensuring such constancy at a pre-processing stage. The present work builds on these previous studies to accomplish the next step, namely *landmark characterization to support subsequent recognition* under different illumination conditions and viewpoints.

II. Saliency detection based on color contrast

A region in an image is considered *salient* if it ranks high in a given feature and its surround ranks high in the opposite feature. The color features considered are based on the opponent colors proposed by Hering [9].

From the input image, Gaussian pyramids corresponding to intensity, orientation and color opponency images are constructed, each with eight spatial scales. A pixel at a fine scale corresponds to a center region, whereas the respective pixel at a coarser scale corresponds to its surround. This multiscale approach is advantageous in that it permits extracting landmarks of varied sizes.

Three sets of partial saliency maps are constructed, corresponding to the intensity, color and orientation features. The partial saliency maps should be combined to obtain one global saliency map. They cannot simply be added, because salient regions present in only a few maps can be masked by noise or less salient regions present in a larger number of maps. The process of combining the partial saliency maps is structured in two stages. In the first stage, the partial saliency maps are normalized by the maximum saliency value obtained at all center-surround scales. In the second stage, the maps are weighted by their information content. The information content of an image is based on their zero-order entropy [11]. Finally, the partial saliency maps are subject to exponentiation and added to compose the global saliency map.

The modifications introduced to the original visual saliency algorithm [7], to improve the color constancy properties, resulted in the *color-ratio visual saliency* algorithm [14], described next.

With the purpose of obtaining contour images with good color constancy properties, Gevers and Smeulders [5] developed a color space based on the color ratio between neighboring pixels. This differential version of color constancy gave us the idea of generalizing the concept of gradient between neighboring pixels to that of center-surround opposition. Thus, invariance of color gradients would turn into the desired invariance of center-surround oppositions. Under this approach, one pixel is replaced by the center region and the other pixel by the surround region. Moreover, the ratios no longer relate color bands, but color opponents, as follows:

$$RG = R_o^c G_o^s / G_o^c R_o^s \tag{1}$$

$$GR = R_o^s G_o^c / G_o^s R_o^c \tag{2}$$

where R_o^c and G_o^c are opponent red and green components at center regions and R_o^s and G_o^s are opponent red and green at surround regions. The same is valid for the yellow and blue components. According to the unichromatic reflection model, assuming that center and surround regions have a locally constant illuminant, the same surface normal and uniform albedo, and the use of narrow-band sensors, we have [14]:

$$C = m_b(\vec{n}, \vec{s}) e(\lambda_c) c_b(\lambda_c) \tag{3}$$

where C is the light sensor response corresponding to a surface patch illuminated by an incident light $e(\lambda)$, λ is the light wavelength, m_b is the body geometric dependency, $\vec{n}$ is the surface normal, $\vec{s}$ is the direction of illumination source, and $c_b(\lambda)$ is the body spectral reflection property. Combining (3) and (1), we have:

$$RG = \frac{(m_b^c(\vec{n},\vec{s})\ e^c(\lambda_R)\ c_b^c(\lambda_R))(m_b^s(\vec{n},\vec{s})\ e^s(\lambda_G)\ c_b^s(\lambda_G))}{(m_b^s(\vec{n},\vec{s})\ e^s(\lambda_R)\ c_b^s(\lambda_R))(m_b^c(\vec{n},\vec{s})\ e^c(\lambda_G)\ c_b^c(\lambda_G))} = \frac{c_b^c(\lambda_R)c_b^s(\lambda_G)}{c_b^s(\lambda_R)c_b^c(\lambda_G)} \quad (\mathbf{4})$$

which is only dependent on the sensors and the surface albedo. The same can be done for Equation (2) and the blue-yellow components. A key feature of these color ratios is their invariance to both intensity and color normalizations, which makes them intrinsically invariant to lighting intensity and illumination color changes. The ratios have a local nature, avoiding the distorting effects possibly introduced by global normalizations. The logarithmic spaces (R_o/G_o) and (Y_o/B_o) permit the computation of the ratio opponencies by simple differences of logarithms across the scales.

III. Delimiting Landmark Regions

Since the extracted salient regions are not necessarily bounded by well-defined contours, nor associated to single elements in the scenes, a refinement step is necessary in the process of determining the boundaries of landmark candidates. As an initial approximation (Figure 1), a minimal rectangular bounding box (Figure 2) is computed for each segmented saliency spot. The objective of the next two processing steps is to get a better fitting of the bounding boxes to the salient features.

In the next step, the colors appearing in each saliency-selected region are identified, and a corresponding *backprojection map* is built, emphasizing where the same colors appear in the whole image. This is performed using histogram backprojection [12].

After this, the size and position of all bounding boxes are adjusted (Figure 2), taking into account the color feature spatial distribution and the respective visual saliency. This is achieved using the *continuously adaptive mean shift* algorithm [2]. This is a non-parametric technique that climbs the gradient of a probability distribution to find the nearest dominant mode, with the capability to adapt the window size. To increase the amount of information associated with the bounding boxes, their immediate surrounding region is also analyzed (Figure 2), giving additional context information to the recognition process.

IV. Landmark Characterization

After the determination of the bounding boxes, region descriptors are extracted. These descriptors should be appropriate to characterize the bounding boxes as signatures of the landmarks and should make the comparison between them possible. Color has proven to be the most suitable of the considered low-level features for outdoor

unstructured environments, where most objects have deformable shapes. The way color features are represented and color descriptions are compared using the adopted representation are described below.

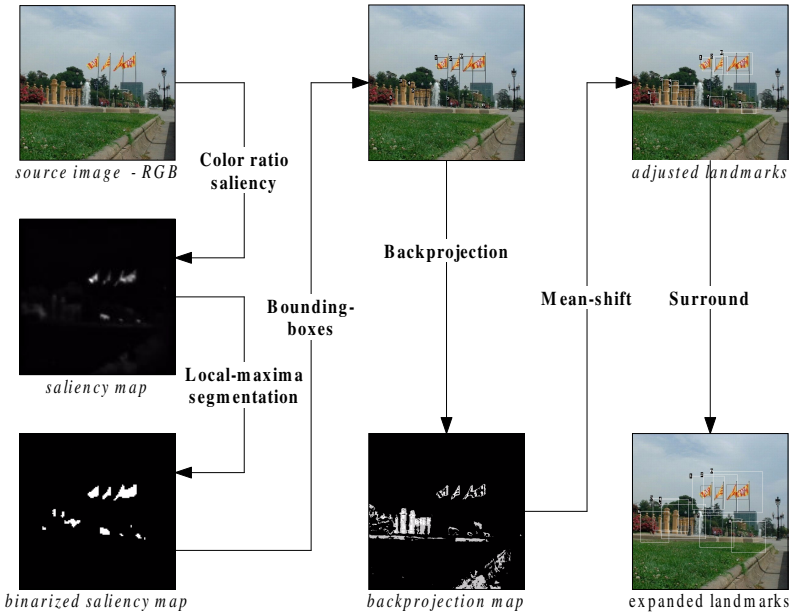

Fig. 1. The process of delimiting the landmark regions. From the source image a saliency map is computed, then this map is segmented, generating the seeds of the landmark regions. These seeds are enclosed by bounding boxes, which are adjusted to the salient elements in the image using color histogram backprojection and mean-shift algorithms. Finally, the landmark bounding boxes are expanded, encompassing the immediate surrounding regions.

Fig. 2. Initial (left), adjusted (center) and expanded (right) landmark bounding boxes.

The most common representation of color in image retrieval and recognition is the color histogram, which captures the global color distribution in an image or region [12, 6]. They are simple to compute and have the properties of invariance to translation, invariance to rotation about an axis perpendicular to the image, and they change smoothly with rotation about other axes, occlusion, and variations in scale. In order to

remove the dependency on the number of pixels that comprise the histogram by comparing histograms of images of different sizes, the histogram can be normalized by dividing each bin count by the total number of pixels. The normalized histogram corresponds to a color probability distribution function.

Taking this considerations into account, the following descriptors to characterize the landmarks were proposed:

1. Normalized chromaticity histogram of salient region inside bounding box.
2. Normalized chromaticity histogram of adjusted bounding box.
3. Normalized chromaticity histogram of expanded bounding box.
4. Normalized saliency histogram of adjusted bounding box
5. Mean saliency of adjusted bounding box.

V. Landmark Matching

Once the feature representation has been defined as a histogram space, the similarity between two images or regions *i* and *j* is described as the distance between their corresponding points h_i and h_j in the histogram space [12].

There are several metrics to evaluate histogram distances. The most common are histogram intersection and Minkowski distances [12]. These distance metrics are quick to compute, but they only compare corresponding bins of the two histograms, disregarding any kind of similarity between colors. This characteristic makes these distance metrics strongly sensitive to slight changes in the distributions. In contrast with Minkowski and intersection distances, the quadratic form metric allows for similarity matching between different colors, and it is defined as follows [6]:

$$d^2_{hist}(h_1, h_2) = (h_1 - h_2)^T \mathbf{A} \ (h_1 - h_2) \tag{5}$$

where h_1 and h_2 are *N*-dimensional color histograms, and $\mathbf{A}$ is the similarity matrix, whose elements a_{ij} denote similarity between bins *i* and *j*. The similarity of landmarks is evaluated with quadratic-form distance by combining the distances between each of the three color histograms stored in the landmark representation. The distances are combined using the root of the sum of the three squared distances.

VI. Experimental Results

From eleven sample scenes in outdoors, 68 landmarks were extracted. To evaluate the retrieval performance of the system, each landmark was taken out of the database, and matched against all other landmarks. Then, the distances to all other landmarks were sorted in ascending order. In image retrieval systems, the quality of matching is usually qualified in terms of *recall* and *precision* figures [4]. Recall is defined as the ratio between the number of relevant images retrieved and the number of all relevant

images in the database. Precision is defined as the ratio between the number of relevant images retrieved and the number of retrieved images.

$$Recall = C_K / M , \qquad Precision = C_K / K \tag{6}$$

where K is the number of retrievals, C_K is the number of relevant matches among all the K retrievals, and M is the number of total number of relevant matches in the database. Another metrics used to quantify the performance of a retrieval system is the *success of target search* index (*STS*). It measures the rank of the first retrieved relevant image (target) in the database with respect to the query, defined as [10]:

$$STS = \left(1 - \frac{rank - 1}{N - 1} \right) \tag{7}$$

where *rank* is the retrieval position of the first retrieved image, and N is the number of images in the database.

The *recall* score (Table 1) obtained was acceptable, considering that the recognition was based solely on color distribution information. This *recall* score indicates few false negative errors. Also the rank of the first (best) retrieved similar landmark was very significant, with the *STS* score near one. The *precision* score obtained is low, indicating the presence of false positives in the retrieval process. This occurs due to the similar color distributions of some detected salient features in different scenes, and since histograms do not provide spatial information about their arrangement, very different images can have similar color distributions, that could mislead into false evaluation of their dissimilarity.

The combined distance form (squared sum of the three region type distances) improves significantly the recall and precision metrics, because of the union of saliency-oriented information with surround information.

Table 1. Recall, STS and precision for the described landmark matching experiment. Resultant measures are shown for each region type individually, and then for a combined form of them.

	Recall	STS	Precision
Spot of saliency bounding box	0.62	0.98	0.24
Adjusted bounding box	0.60	0.99	0.21
Expanded bounding box	0.53	0.98	0.17
Combined histograms	0.70	0.99	0.26

The computational time of the main tasks (Table 2) were evaluated using a standard PC computer (Pentium III 900MHz, 256Mb DRAM, Microsoft Windows XP). It can be observed that the saliency detection is the task that demands more computational time, and that the histograms are computed very quickly. In the landmark comparison phase, although the quadratic-form histogram distances could take a lot of

time to be computed, the small size of the histograms (16x16 bins) keeps computational time low for this task.

Table 2. Computational complexity and execution times of the main tasks related to landmark characterization and matching. N is the number of pixels in the input image and M is the number of bins in the histograms. Data is shown with two significant digits.

Task	Computational complexity	Seconds
Visual saliency with color ratios (512x512 pixels)	O(*N*)	0.81
256-bin histogram (16x16 bins)	O(*N*)	0.00015
Landmark characterization	O(*N*+*M*)	0.039
Quadratic-form histogram distance	O(*M*)	0.0085
Landmark matching	O(*M*)	0.028

VII. Discussion

In a pioneering work on image indexing, Swain and Ballard [12] pointed out that, for real-time object recognition, color-based algorithms were especially promising, due to their fast performance and their capability to deal with viewpoint changes, object deformations, and inaccurate segmentation. They considered a challenging problem to identify the region from which to extract the histogram to be used as object signature for recognition purposes.

This is exactly the first contribution of the current research, proposing a novel saliency detection algorithm with embedded color constancy properties, and using this information to identify and delimit image regions that can be used as landmarks.

A second contribution is the landmark characterization that, going beyond the single histogram, combines saliency and chromaticity into a robust and stable signature, as confirmed by experimentation.

Indeed, the results show good recognition performance, in terms of both recall and rank indices. However, the discrimination between landmarks requires improvement to avoid false positive mistakes, i.e., retrieving landmarks from the database that do not correspond to the query landmark. This shortcoming is not a critical one in our application, since a rough knowledge of the robot trajectory can help to disambiguate between landmarks with similar appearance.

Acknowledgments

The authors would like to thank Enric Celaya and Pablo Jimenez for productive discussions about visual saliency and robot localization. This work is partially supported by the Spanish Council of Science and Technology under project “Vision-based re-

configurable navigation system for legged and wheeled robots in natural environments" (DPI 2003-5193).

References

[1] J. Batlle, A. Casals, J. Freixenet, and J. Martí, "A review on strategies for recognizing natural objects in colour images of outdoor scenes," *Image and Vision Computing*, vol. 18, pp. 515-530, 2000.

[2] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," Fourth IEEE Workshop on Applications of Computer Vision, pp. 214-219, 1998.

[3] W. Burgard, A. Derr, D. Fox, and A. B. Cremers, "Integrating global position estimation and position tracking for mobile robots: the dynamic Markov localization approach," IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS '98), Canada, pp. 730-735, 1998.

[4] Y. Deng, B. S. Manjunath, C. Kenney, M. S. Moore, and H. Shin, "An efficient color representation for image retrieval," *IEEE Trans.on Image Processing*, vol. 10, pp. 140-147, 2001.

[5] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, pp. 453-464, 1999.

[6] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, "Efficient color histogram indexing for quadratic form distance functions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, pp. 729-736, 1995.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259, 1998.

[8] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.

[9] S. J. Sangwine and R. E. N. Horne, *The color image processing handbook*, 1st ed. London: Chapman & Hall, 1998.

[10] R. Schettini, G. Ciocca, and S. Zuffi, "A survey of methods for colour image indexing and retrieval in image databases," in *Color Imaging Science: Exploiting Digital Media*, M. R. Luo and L. MacDonald, Eds., 1st ed: John Wiley & Sons, 2002, pp. 183-211.

[11] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 1948.

[12] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11-32, 1991.

[13] E. Todt and C. Torras, "Detection of natural landmarks through multiscale opponent features," 15th International Conference on Pattern Recognition, Barcelona, Spain, pp. 976 - 979, 2000.

[14] E. Todt and C. Torras, "Detecting salient cues through illumination-invariant color ratios," *Robotics and Autonomous Systems*, vol. 48, pp. 111-130, 2004.