Bayesian supervised segmentation of objects in natural images using low-level information

Jiří Boldyš
Helsinki University of Technology
Laboratory of Media Technology
PL 6400, 02015 TKK, Espoo, Finland
Jiri.Boldys@hut.fi

Abstract

Detection of particular meaningful objects in images is of great importance in many fields, including image retrieval or image quality analysis. In this contribution, eleven frequently observed objects (areas) in natural images are learned and detected. The presented algorithm is based on region merging and Bayesian decision theory. The main goal is not perfect recognition, since for that purpose it is necessary to use higher-level knowledge about the image content. Merging of segments proceeds only up to a reliable point, not to merge different categories. Unique merging criteria combine the values of probabilities attached to the segments for all the most likely categories, color and texture features and information about edges. Results are demonstrated on a few images and discussed.

1 Introduction

In the last few years, demands on the level of image analysis have significantly increased. More often, particular objects or areas contained in the image are recognized, and the semantic description is derived. The results of such image interpretation can be used in different fields.

The most apparent application is image retrieval. Higher-level image descriptors better correspond to queries usually formulated by a user. For a review of content-based image retrieval methods, see e.g. [9]. The work [3] reviews image retrieval by semantic content.

Another possible application of image analysis is image enhancement and image quality. Take, for example, an image which contains sky. It is then possible to de-noise this area according to assumed smoothness or to change its color so that it would be visually more pleasing. This task is closely connected to image quality analysis, where the notions of genuiness and naturalness can be evaluated [13].

There has been two main approaches to image analysis. The first approach is trying to employ low-level features and establish their correspondence to a higher-level image description, as in [14]. In [12], low-level features are used to calculate semantically meaningful classification of vacation images.

The second approach employs the knowledge about the objects characteristics and about their qualitative and quantitative relationships. It attempts to describe action going on or even feelings about the image. For a survey of image understanding systems based on higher-level knowledge, see [1].

Objective of this contribution is use of low-level features for supervised recognition of objects in images of natural scenes and their simultaneous segmentation. Since perfect segmentation using only low-level features is impossible, the segmentation proceeds as far as using only low-level features is reliable. Output of the algorithm is then a set of segments where for each segment, the probabilities are calculated that the segment belongs to the categories of interest here. The output is then prepared for later use of higher-level knowledge about the segments relationships, see e.g. [10] or [5].

This paper is focused on natural image analysis. Eleven categories are learned and detected, which are listed below. However, the algorithm should reasonably work on all kinds of images, containing even unknown objects.

The algorithm relies on segment merging with use of a unique combination of criteria, based on the probabilities of the most likely categories, on color, texture and edge information. The details are given below, in Section 2. The segmentation results are presented in Section 3, together with a thorough discussion.

1.1 Categories

The set of categories to be detected in natural images can be chosen according to different criteria. Taking into account all the possible applications of the segmentation algorithm mentioned in the introduction, the most appealing criterion is probably the frequency of their occurrence in typical images.

In this paper, the following set of eleven categories was detected: { sky, grass, water, sand, soil, reed, tree-top, road, stones, mountains, snow }.

Due to the probabilistic nature of the algorithm, the distinction between all the categories does not have to be sharp. The detected categories are rather offered as options for a certain segment and the final decision often heavily depends on the image contents.

Thus, for example, when learning the category grass, distinction was not made between green, yellow or brown areas, if they were reasonably small and the overall area was really perceived as a grass field. On the other hand, homogeneously dry grass field is contained in the category reed, due to their perceptual proximity.

The proposed set roughly corresponds to perceptual differences observed when visually examining a natural image database. For example, the category mountains could be described only by the other categories snow, stones, grass and tree-top. However, the first impression, when observing an image with mountains, does not correspond to any of the other categories. This aspect is especially important in the field of image retrieval.

Another example of a disputable division is pair of categories soil and sand, because there is only little difference between them. The concept of this paper is rather to detect all possible categories for one particular segment. Therefore, if the color of a sand segment reminds rather soil, it would be detected as sand only as the second option. However, for higher-level image analysis it is then probably easier to work with the category sand, for example in a seacoast scene.

2 Description of the algorithm

Ability of an algorithm to distinguish between different categories depends on the set of features used for their description. In applications built to recognize certain objects precisely, sophisticated models can be used, see e.g. [7], where a physical model of sky is built. However, the segmentation performed in the algorithm proposed here is rather supposed to support the other algorithms using higher-level knowledge, although as good as possible discriminability is also of value. Therefore, the feature set used here can be relatively simple.

Other reason for keeping the features simple is a large variability of texture of the categories of interest here. For example, grass can be very rough and also smooth, depending on the distance of the camera. The color of grass can also vary from green to yellow or brown but it is not always desirable to recognize the spots other than green as dry grass or soil. Thus, only simple wavelet features are used to describe texture, rather on the principle that certain feature value is unusual for a given category.

2.1 Features

Probably the most important characteristic for the natural image description is color. The CIELAB [4] color coordinate system is used because it resembles human color perception and the segmentation results should be perceptually acceptable. Since the algorithm behaves robustly, the transformation to the CIELAB coordinates needs only be carried out approximately, assuming standard display phosphors and the white point D_{65}. Three features, the color coordinates L^*, a^*, and b^* are calculated for every pixel of the image.

Since for evaluation of the probabilities only the coordinates a^* and b^* are calculated, slight invariance to varying illumination conditions is achieved. On the other hand, it is more difficult to distinguish water from the sky, or even from the tree-top. These problems are to be resolved using higher-level knowledge. Information about the lightness L^* is used only during segmentation.

The second set of features was chosen to approximately describe roughness of the texture at different scales. Texture descriptors cannot be precisely local, therefore we divide the image into non-overlapping blocks and the texture features are attached to the blocks.

Wavelet transform and wavelet packet coefficients [11] are increasingly used for texture segmentation [6]. Magnitudes of the Haar wavelet transform coefficients for the detail bands (horizontal, vertical, diagonal) of the first three levels were used in this paper. 48 first level, 12 second level, and 3 third level coefficients were attached to every block of size 8×8 pixels. Haar wavelets were chosen because of their short support, which means that they form a block non-overlapping basis.

2.2 Bayesian framework

The eleven categories are recognized based on the Bayes formula [2]. Having a particular segment A, the probability that the segment is from the category $\omega_{A,c}$ is

$$P(\omega_{A,c}|\mathbf{x}_A) = \frac{P(\mathbf{x}_A|\omega_{A,c})P(\omega_{A,c})}{P(\mathbf{x}_A)}. \quad (1)$$

We are working with discrete probabilities, that a feature value falls to a certain bin in a discretized feature space. $\mathbf{x}_A$ is a feature vector calculated on the pixels of the area A. The second index c in $\omega_{A,c}$ denotes category, the first index denotes the area (neighborhood, pixel) on which the probability is evaluated. $P(\mathbf{x}_A|\omega_{A,c})$ is the likelihood of measuring the feature vector $\mathbf{x}_A$ when the area A belongs to the category $\omega_{A,c}$. $P(\omega_{A,c})$ is an a priori probability of detecting the category $\omega_{A,c}$ and $P(\mathbf{x}_A)$ is the probability of measuring the feature vector $\mathbf{x}_A$.

In accordance to the minimum-error-rate classification, segment A is from the category $\omega_{A,i}$, if

$$P(\omega_{A,i}|\mathbf{x}_A) > P(\omega_{A,j}|\mathbf{x}_A)$$

for all $j \neq i$.

Independence of the color $\mathbf{ab}_A$ and texture $\mathbf{w}_A$ features is assumed. Therefore,

$$P(\mathbf{x}_A|\omega_{A,c}) = P(\mathbf{ab}_A|\omega_{A,c})\, P(\mathbf{w}_A|\omega_{A,c}).$$

Further, though unrealistically, pixel-wise independence of the features is assumed. Thus, for the color features

$$P(\mathbf{ab}_A|\omega_{A,c}) = \prod_p P((a,b)_p|\omega_{p,c}),$$

where p goes over all pixels of the area A.

The wavelet features calculated on different blocks of 8×8 pixels are also assumed independent:

$$P(\mathbf{w}_A|\omega_{A,c}) = \prod_s P(\mathbf{w}_{8\times 8,s}|\omega_{8\times 8,c}).$$

s goes over all 8×8 blocks of the area A. The feature vector $\mathbf{w}_{8\times 8,s}$ consists of absolute values of the detail wavelet coefficients for 3 levels. The probability $P(\mathbf{w}_{8\times 8,s}|\omega_{8\times 8,c})$ is again a product of probabilities to measure these (supposedly) independent features for the category ω_c on a proper neighborhood.

In this paper, a priori information about the categories is not used for classification. The segments are classified only according to their color and texture properties and not according to statistics of occurrence of the categories. As the discriminant functions, log-likelihood from (1) is used:

$$\log P(\mathbf{x}_A|\omega_{A,c}) = \log P(\mathbf{ab}_A|\omega_{A,c}) + \log P(\mathbf{w}_A|\omega_{A,c}) \quad (2)$$

with

$$\log P(\mathbf{ab}_A|\omega_{A,c}) = \sum_p \log P((a,b)_p|\omega_{p,c})$$

and analogously with the second term. Although values of the discriminant functions (2) are always used during the calculation, they are further, for simplicity, called probabilities.

Given the area (segment) A contains s blocks 8×8. It is desirable to be able to compare the probabilities (2) for segments of different sizes. Therefore, they are normalized to the number of blocks covered by the segment A. Moreover, just to get observable numbers, the probabilities are normalized by one more factor, equal to the number of features evaluated at one block.

2.3 Database for Supervised Learning

A medium-size (500) database of Corel images was used for learning of the categories. In all images, all eleven categories were manually segmented and labelled.

2.4 Processing of the Feature Histograms

The histograms for every feature ($2-$D histogram in case of a^* and b^* coordinates) and for every category were normalized to get the probability that a particular feature has a certain value. In case, the set of images used for learning is not representative enough, two precautions were made. Every probability was median-filtered (with symmetric border reflection) first. Secondly, for the values where the probability is negligible or zero, a small constant was added. This precaution increases robustness of the algorithm.

2.5 Merging

After a long experimentation phase, the following merging scheme showed up to work best (regarding the objectives) and robustly, without being over-sensitive to change of parameters.

Merging consists of four steps. To describe them efficiently, expression 'color criterion' will be used as a synonym for average L^*, a^*, and b^* values for every block. 'Roughness criterion' stands for the sum over all three bands (horizontal, vertical, diagonal) of average magnitudes of wavelet coefficients for the first level of decomposition. Expression 'probability criterion' means that the probabilities of two neighboring segments for the most likely categories were compared. The probability criterion was imposed depending on the merging step. It prevents merging of, for example, sky and water categories in case, there is a smooth transition between them in terms of only color and roughness criteria.

In the first step, very conservative and fast initial segmentation was performed. The probability, color and roughness criteria were used separately at smooth and rough image areas. The division to smooth and rough is determined by thresholding the roughness criterion and the initial segmentation was not allowed to merge across them.

After the first stage, region adjacency graph [8] was constructed and used in the next stages. The second stage is analogous to the first stage, but the thresholds are relaxed. The algorithm merges always the closest (according to criteria) neighbor. In the third stage, segments can merge across the smooth and rough areas. Merging of smaller segments is encouraged.

After the third stage, some of the smooth segments are still not merged. This is corrected in the last stage, which merges two segments according to magnitude and relative length of edge between them. Color criterion (without lightness) and probability criterion are added for the rough segments.

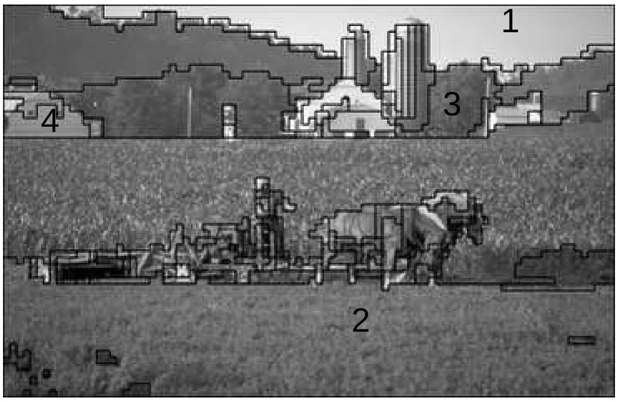
3 Results and Discussion

The proposed algorithm was tested on many images not contained in the database, which was used for learning. All the results were checked visually by three observers and were found acceptable, having in mind the objective of the algorithm. The artifacts found in the results are described by means of the examples below. In this section, all the images are presented as gray-level, although the experiments were performed on color images.

One typical example of the results is shown in Fig. 1. In the top sub-image, the original image is histogram-equalized to make low-contrast details and areas better visible. In the bottom sub-image, there is original image with overlaid contours delineating the resulting segments.

Due to the nature of the algorithm, the results are to a certain degree stochastic. However, the number of cases when

(a) Histogram-equalized original



(b) Segmentation

Figure 1. Example 1

two different categories are merged is very low. The segments visually correspond to areas with different color and texture. Nevertheless, there are also frequent cases when, for example, the sky or grass is divided into more segments, although visually the area is homogeneous. It is sometimes caused by too large a difference of probabilities on these segments, calculated for the most likely categories.

Sometimes a small segment covers an image area, which at the first look does not seem to be different so much from the surrounding segment. However, objectively, there is always one criterion which prevents merging them.

For example, there are small segments in the middle of the (lower) grass field in Fig. 1. Particularly in this case, additional criteria could have been developed, taking into account that it is a small inner segment of a big surrounding segment and in such a case the thresholds could be relaxed a little. However, the main goal of this paper is to pre-segment the image for a higher-level image understanding algorithm, and for this purpose, little over-segmentation is not a problem.

Sometimes there are two rough-surface regions which visually look the same but they are not merged. It happens when there is accidentally an edge along the boundary between these segments.

Table 1. First three choices for the segments marked in Fig. 1

seg.	choice 1	choice 2	choice 3
1	sky (-1.37)	water (-1.54)	sand (-1.85)
2	grass (-1.60)	tree-top (-2.13)	water (-2.14)
3	water (-2.09)	tree-top (-2.15)	grass (-2.25)
4	water (-2.00)	tree-top (-2.17)	mount. (-2.19)

Since the algorithm starts with dividing the image into 8×8 blocks, it effectively averages the texture properties. Thus, it deals well with very rugged areas where, for example, color is changing locally significantly. On the other hand, it sometimes creates segments containing two categories, for example the segments covering the border between forest and sky in Fig. 1. This is also the reason, why the presented method over-segments close-up images.

Generally speaking, it is better to set up thresholds conservatively, than to cause false merges. A few cases when two segments should merge but are separate, do not cause any problems.

A list of segments is an output of the segmentation algorithm. Together with every segment, there is a sorted list of categories, according to their probability (up to some threshold).

In Table 1, three most probable categories for selected segments of the image from Fig. 1 are presented. The selected segments are denoted by numbers, inscribed in the image. Choice 1 denotes the most likely category. Every possible category is accompanied by its probability.

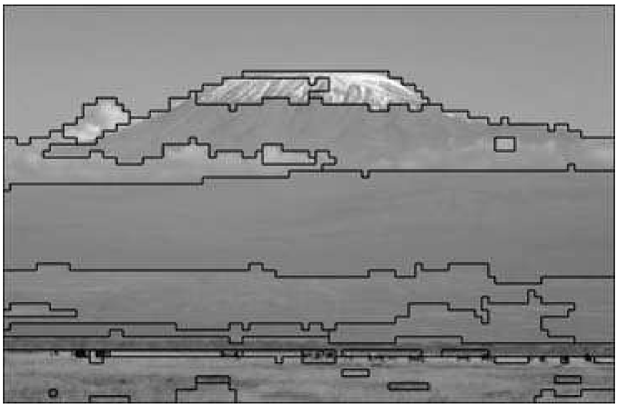
Often the most likely category is not correctly detected or there are many categories with similar high probabilities. It is not a serious obstacle, since this algorithm should only pre-segment the image. Certain ambiguity or incorrect classification can be then resolved by imposing spatial relation constrains.

This approach was motivated by critical survey of many natural images. It was found out that many areas in natural images are perceptually very similar, if we focus only on them. Humans can reliably recognize the sky and water, or soil and sand, because they see the context. Therefore, the features are kept simple and are not aimed to be capable of a perfect recognition. For example, the mountain category is problematic in this sense.

The image from Fig. 2 contains some smooth transitions between different areas and transitions, noticeable on the non-enhanced image only with difficulties. The upper-left cloud is not merged with the rest of the sky, because these two segments have substantially different probability of being water, as the second choice for the sky segment. Although this criterion might look unwanted, it is very useful in other cases, where the transition between two categories is smooth in both color and texture.

In Fig. 3, there is a non-natural object, which was not learned. The algorithm then decides for the closest of the

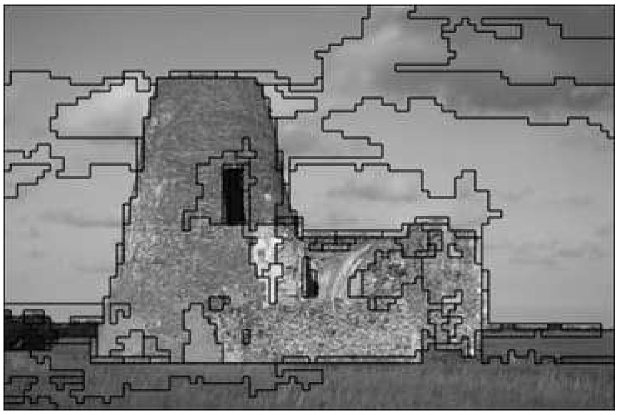
(a) Histogram-equalized original



(b) Segmentation

Figure 2. Example 2

(a) Histogram-equalized original



(b) Segmentation

Figure 3. Example 3

learned categories. If the object does not resemble any learned category (determined by threshold), the algorithm is merges according to chromaticity coordinates and roughness only.

The last example is presented in Fig. 4. There is one segment between the left and middle tree, which contains too large area. However, looking at it without knowledge of the image context, it is not perceived heterogeneous.

In certain applications, where the segment contours should be as precise as possible, it is not difficult to refine them, as shown in the bottom sub-image of Fig. 4. Beginning with the largest segments, their borders were relaxed according to color distance to the neighboring segments.

Segmentation of one image of usual size (500×700 pixels) takes over one minute on a 1.1 GHz computer. The time depends heavily on quality of the initial segmentation. This is not a severe limitation for off-line image analysis. However, speeding up the algorithm is one of the future directions of this research.

As was already mentioned, applicability of the presented algorithm is lower on the close-up images. The segmentation might not resemble visual perception, although merges across categories are, in principle, prevented.

Another problematic group of scenes is that with unusual illumination. The algorithm was trained on images in normal illumination conditions. There is certain invariance to illumination changes due to use of only a^* and b^* color coordinates, when calculating the probabilities and due to large diversity of the training data. However, the algorithm fails in segmentation of late-evening or sunset images.

4 Conclusions

The algorithm proposed in this paper proved to be powerful and robust in detection of objects in images of natural scenes. Moreover, the segmentation results are visually pleasing. Probabilistic description of the segments can serve as a promising input to higher-level image analysis methods, for example to probability relaxation method.

The next plans are to improve speed of the algorithm and to experiment with employment of the higher-level methods.

5 Acknowledgments

This work has been financed by the Graphic Industry Research Foundation (GTTS), Helsinki, Finland.

The author would like to thank to Jyri Kivinen for many inspiring discussions and helpful comments.

The author is currently on leave, from the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic.

References

[1] D. Crevier and R. Lepage. Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding*, 67(2):161–185, August 1997.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons Inc., 2001.

[3] J. P. Eakins. Towards intelligent image retrieval. *Pattern Recognition*, 35:3–14, 2002.

[4] M. Fairchild. *Color Appearance Models*. Addison-Wesley, Reading, MA, 1998.

[5] J. A. Feldman and Y. Yakimovsky. Decision theory and artificial intelligence: I. a semantics-based region analyzer. *Artificial Intelligence*, 5:349–371, January 1974.

[6] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1190, November 1993.

[7] J. Luo and S. P. Etz. A physical model-based approach to detecting sky in photographic images. *IEEE Transactions on Image Processing*, 11(3):201–212, March 2002.

[8] T. Pavlidis. *Structural Pattern Recognition*. Springer-Verlag, New York, 1977.

[9] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.

[10] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Brooks/Cole, Pacific Grove, CA, 1999.

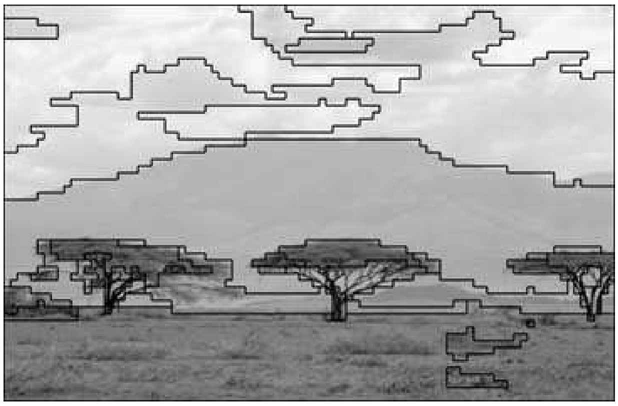
[11] G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, USA, 1996.

[12] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 10(1):117–130, January 2001.

[13] S. Yendrikhovskij. Image quality: Between science and fiction. *IS&Ts PICS Conference*, pages 173–178, April 1999.

[14] R. Zhao and W. I. Grosky. Negotiating the semantic gap: from feature maps to semantic landscapes. *Pattern Recognition*, 35(3):593–600, 2002.

(a) Histogram-equalized original



(b) Segmentation

(c) Segmentation with refined borders

Figure 4. Example 4