MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
3-30

# Building Classification of Terrestrial Images by Generic Geometric Hierarchical Cluster Analysis Features

Gerd Brunner and Hans Burkhardt
Institute for Pattern Recognition and Image Processing,
Computer Science Department, University of Freiburg,
Georges-Koehler-Allee 052,
79110 Freiburg, Germany
E-mail: {gbrunner, Hans.Burkhardt}@informatik.uni-freiburg.de

## Abstract

*The scope of this paper is the challenging task of classifying terrestrial images of buildings, automatically. Straight line segments and their connectivity incorporate significant information about object shapes. Man-made buildings exhibit special generic shapes which are extracted from embedded spatial and angular line segment relationships by cluster analysis. After employing an agglomerative hierarchical cluster analysis we obtain geometrical structure information features on different scales. For the classification process we apply support vector machines (SVM) with polynomial and radial basis function (RBF) kernels to separate the feature space by a hyperplane into 2 classes. The method is applied to an image collection taken from the Corel image database and compared with traditional edge-orientation histogram features. We obtained a 88 % true positive classification rate (recall) with an F-measure value of 81.3 %.*

## 1 Introduction

Nowadays, we observe a permanent increase of image data, resulting in a demand of qualitative and quantitative image retrieval and classification systems. Although, many researchers have devoted much time to the area of image classification, it remains in general an open and challenging area.

In this paper, we deal with the hard problem of classifying images of buildings. The semantic concept of how buildings look like is not easy to define, since there exists no common shape, size, appearance or color of a building. Moreover, it is not entirely clear which objects do not belong anymore to the class of buildings, as illustrated by the example of memorials or fountains. The problem of classifying buildings has already been applied to aerial imagery data [5] [6], exhibiting a slight different formulation of the task than in our case, since only projections of building roofs are visible from satellites or air-planes.

We are dealing with terrestrial images classification of buildings [3] [7], featuring higher spatial resolution. To approach a solution of the building classification problem, 2 steps are needed. In the first step, proper shape features must be extracted and in the second step a classifier has to be designed and adjusted to the feature space. Considering the first part, [8] applied perceptual grouping of L and U-junctions to edges for describing shapes of man-made objects. The second step consists of creating a discriminative classifier.

Our approach employs the extraction of generic shape features, obtained by agglomerative hierarchical cluster analysis. As classifier we have decided for support vector machines (SVM), due to their great generalization abilities. We utilize polynomial and RBF kernels for the classification of building versus non-building of approximately 2100 images taken from the Corel image database [11] [15]. The proposed method will be compared with the well established edge-orientation histogram [4] features.

## 2 Feature Extraction

We decided to use line segment features since they contain important shape information of man-made objects and especially of buildings. In the following we describe how the line segments are computed.

The basic approach to edge detection is to compute spatial derivatives of an intensity image. The computation of the derivatives is mostly approximated by convolution techniques. We are using the well known Canny edge filter [2], which can be approximated by the derivative of a Gaussian.

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{1}$$

where $x$ and $y$ are the image co-ordinates and $\sigma$ is the standard deviation of the associated probability distribution. The edge location is at the local maximum in the direction $\mathbf{n}$ of the operator; $G_n = \frac{\partial \mathbf{G}}{\partial \mathbf{n}} = \mathbf{n}\nabla\mathbf{G}$, convolved with the image $\mathbf{g}$:

$$\frac{\partial}{\partial \mathbf{n}} G_{\mathbf{n}} * \mathbf{g} = \mathbf{0}. \tag{2}$$

The Canny detector is optimal for step edges corrupted by white noise. To extract salient information from an edge image, further data processing is inevitable. An edge image typically shows many "edge points" and a bunch of very short edge lines, see the second upper image in Figure 1. We perform the edge linking process by tracking

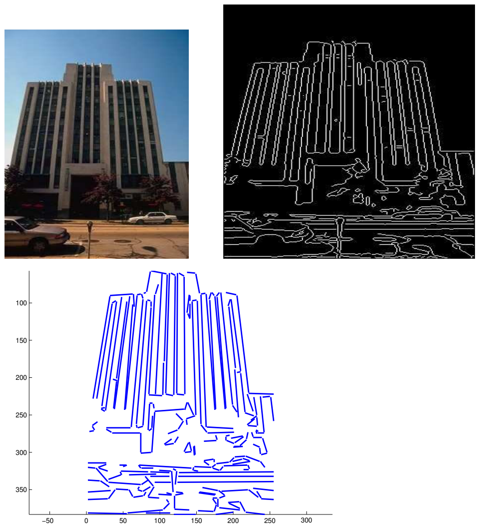

Figure 1: The upper left image is a typical image of the class building. The upper right panel shows the corresponding edge image. The lower image shows the resulting line segments.

edge points and merging broken lines by some variable distance- and angle-tolerance inspired by [12][9]. The left image in the second row of Figure 1 shows the extracted line segments, which are used for further computations.

### 2.1 Hierarchical cluster analysis

Cluster analysis is a method of multivariate statistics to reveal homogenous groups of objects, based on their characteristics. The basic task of cluster analysis is to partition a set $\mathbf{S}$ consisting of $m$ points in an Euclidian space $\mathbf{R}^n$ into $k$ clusters, where each group or clusters should be different from other groups with respect to the same characteristics.

One of the most popular clustering methods is the k-means algorithm, where k is the fixed number of clusters and has to be known at the onset. Since the number of clusters is in our case not known in advance and additionally may vary for different images, we have decided for hierarchical clustering. Moreover, we are able to omit the cluster-center initialization problem, which has a crucial impact on the performance of k-means algorithms. We employ agglomerative hierarchical clustering which takes each entity as a single cluster to start off with and then builds bigger and bigger clusters by grouping similar entities together, until the entire dataset is encapsulated into one final cluster.

Before, applying the cluster analysis we introduce a special weighting schema to cope with the interrelationships of the line segments, their lengths, distances from each other and their relative angles.

$$S^w_{mn} = S_{mn} * W^a_{nm} * W^l_{nm} * W^d_{nm}, \tag{3}$$

where $S_{mn}$ is a similarity matrix containing relative line segment distances and angles. $W^a_{nm}$, $W^l_{nm}$

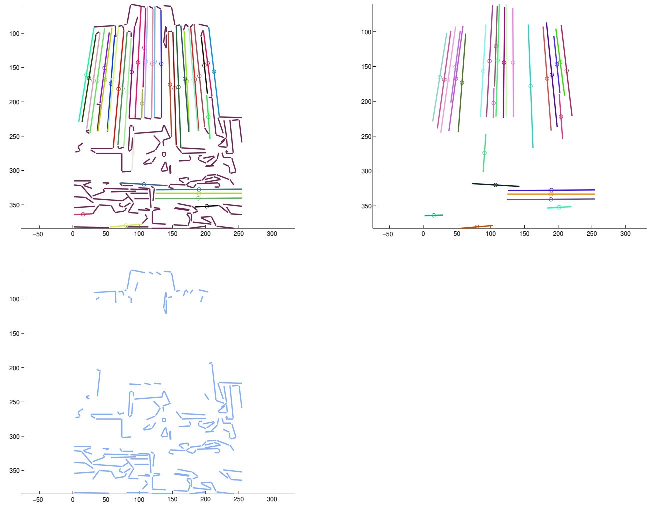

Figure 2: The upper left image shows all resulting cluster drawn with different colors for the image in Figure 1. For a better perception we show in the upper right image, line segments which have been assigned special importance, since they exhibit the geometric structure of the original image. The image in the lower panel shows line segments of lower importance - they might be interpreted as ”noise”.

and $W^d_{nm}$ are weights for the line segment lengths, distances from each other and angles from each other, respectively. The newly formed similarity matrix comprises generic line segment information specifying special relationships of line segments. For describing buildings we give a higher weight to “almost” parallel and perpendicular line segments of longer lengths. Figure 2 shows the resulting clusters after our weighting process.

## 3 Support Vector Machine

Support vector machines have been recently successfully applied to different image classification problems [3] [7]. For a detailed general description of support vector machines we refer to [14].

We will just review some very basics of 2-class SVM’s. Assume, $(x_i, y_i)_{1\leq i \leq N}$ is a set of training patterns, where each pattern $x_i \in \mathbf{R}^n$, with $n$ representing the dimension of the input space and $y_i \in \{-1, 1\}$ are the class labels. The SVM will find hyperplanes that separate the training data by a maximal margin. Thus, all elements located on one side of the hyperplane belong to class one and elements on the other side belong to class two. The so-called *support vectors* are elements of the training set that lie closest to the hyperplane. SVM’s permit us to solve nonlinear decision problems by a kernel-based transformation fulfilling Mercer’s condition [1].

$$K(x_i, x_j) = \Phi(x_i).\Phi(x_j). \tag{4}$$

Thus, a dot product can be computed in a higher dimensional (possibly infinite dimensional) Euclidean space, re-

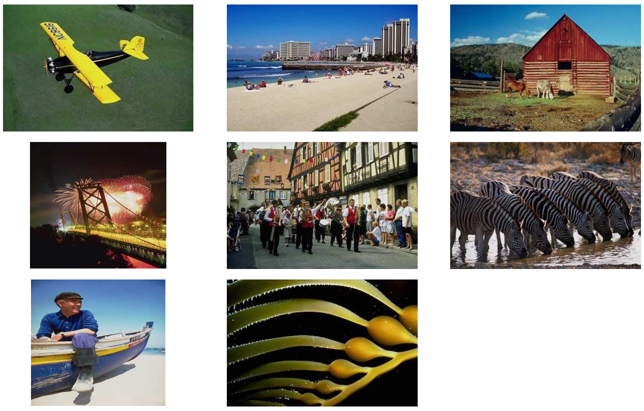

Figure 3: The present figure visualizes the content of the used image database. Due to space limitations we show only 8 random images where each one was chosen from a different class.

sulting in the following decision function

$$f(x) = \sum_{i=1}^{N} (\alpha_i y_i K(x, x_i) + b). \quad (5)$$

We are applying in our experiments the commonly used polynomial kernels of degree $p$, and Gaussian radial basis function (RBF) kernel [13].

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}.\mathbf{y} + 1)^p \quad (6)$$
$$K(\mathbf{x}, \mathbf{y}) = e^{-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}}. \quad (7)$$

# 4 Results

## 4.1 Methodology

We employ the above described approach to classify, whether, an image belongs to the class of buildings or non-buildings. We compare our feature extraction method with the well known and successfully applied edge orientation histogram feature. Both methods are compared with various SVM parameters and kernels.

For our experiments we have taken a subset of the Corel image collection, consisting of almost 2100 images, featuring 21 different classes containing images of the following areas: *buildings, aviation, beaches, earth, leisure, winter, flowers, cars, mountains, cowboys, sunsets, costal, people, underwater, parades, minerals, mammals, jewelry, fireworks, farms and butterfly*. The image class memberships have been decided by Corel. The number of images per class is equal.

Some images are semantically not very consistent with their class labels, e.g. one can find in the class *earth* also images of towns, bridges, mountains or clouds. However, we did not change the original labeling for the sake of comparison. To give the reader an impression about the other image classes we show in Figure 3, 8 random images out of all other classes. Due to lack of space we only show 8 images, where each image is out of a different randomly chosen class.

Table 1: Number of images used for training and testing the SVM.

| Training Images | |
|---|---|
| Class: Building | Class: Non-Building |
| 50 | 39 |

| Testing Images | |
|---|---|
| Class: Building | Class: Non-Building |
| 50 | 2153 |

Table 2: Classification results obtained with edge-orientation histogram features. $\gamma = 0.45$, $C = 2x10^{11}$.

| Measure | SVM-Kernel | |
|---|---|---|
| | RBF [%] | Polynomial [Deg:6] [%] |
| TP | 78.00 | 80.00 |
| F | 72.81 | 73.64 |
| AC | 70.87 | 71.36 |
| FP | 36.26 | 37.28 |
| TN | 63.74 | 62.72 |
| FN | 22.00 | 20.00 |
| P | 68.27 | 68.21 |

## 4.2 Discussion

To validate the obtained results we use usual descriptors obtained from a 2-class confusion matrix which contained the correct classifications of each class and their mismatches. The parameters displayed in Table 2 and Table 3 have the following meanings:

- Recall or true positive rate (TP): Proportion of correctly identified buildings.
- F-measure (F): Measure for the overall performance [10].
- Accuracy (AC): Proportional number of correct predictions.
- False positive (FP): Percentage of incorrect classified non-buildings.
- True negative (TN): Percentage of correct classified non-buildings.
- False negative (FN): Proportional number of incorrect classified buildings.
- Precision (P): Proportion of correct classified buildings.

The classification results using edge-orientation histograms can be seen in Table 2 and the output of our proposed clustered line segments feature histograms in Table 3. Note, that the SVM set-up was chosen identically to clearly show the performance based on different features. Figure 4 displays some classified images of the class building. The results verify a higher discrimination power of our approach over the edge-orientation histogram method.

Table 3: Classification results obtained with clustered line segments histograms. $\gamma = 0.45$, $C = 2x10^{11}$.

| Measure | SVM-Kernel | |
|---|---|---|
| | RBF [%] | Polynomial [Deg:6] [%] |
| TP | 88.0 | 86.00 |
| F | 81.30 | 78.07 |
| AC | 79.76 | 75.84 |
| FP | 28.48 | 34.32 |
| TN | 71.52 | 65.68 |
| FN | 12.00 | 14.00 |
| P | 75.55 | 71.47 |

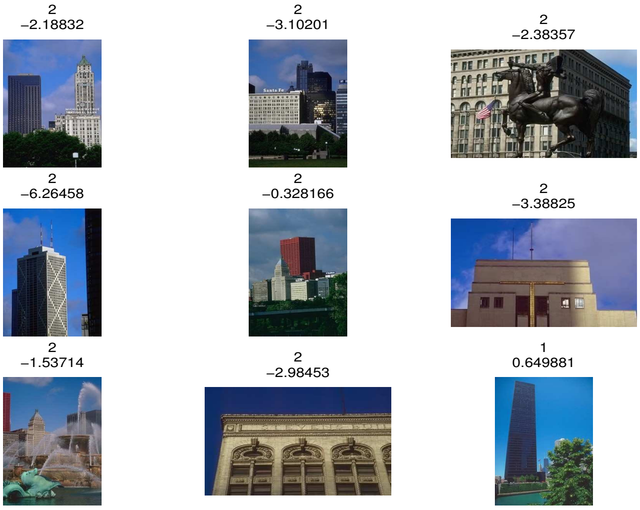


Figure 4: Some classified images of the building class obtained with the RBF kernel. The first number above each image represents the class, where 2 means building and 1 non-building. The second value indicates the distance from the hyperplane. Note, that the last image is wrong classified.

## 5 Conclusions

We have proposed a geometric feature extraction method based on a special weighted hierarchical cluster analysis. The proposed features capture the intrinsic interrelationships of line segments, containing a high discriminative power verified by support vector machines with different kernel functions. We have compared our proposed features with the well established edge-orientation histogram feature. The results proof that our features possess a higher discrimination ability for the class of buildings.

## 6 Acknowledgments

This research was supported by the BMBF I-Search project.

## References

[1] C. J. C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[2] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[3] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. *IEEE Transaction on Neural Networks*, 10(5):1055–1064, 1999.

[4] N. A. Evans A. C., Thacker and J. E. W. Mayhew. The use of geometric histograms for model-based object recognition. In *Proc. BMVC93*, 1993.

[5] A. Fischer, T. H. Kolbe, F. Lang, A. B. Cremers, W. Förstner, L. Plümer, and V. Steinhage. Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D. *Computer Vision and Image Understanding: CVIU*, 72(2):185–203, 1998.

[6] S. B.-M. Gerke M., Heipke C. Building extraction from aerial imagery using a generic scene model and invariant geometric moments. In *Proceedings of the IEEE/ISPRS joint Workshop on Remote Sensing and Data Fusion over Urban Areas, University of Pavia, Rome (Italy)*, pages 85–89, Nov. 2001.

[7] W. HL and C. MU. Image semantic classification by using SVM. *Journal of Software*, 14(11):1891–1899, 2003.

[8] Q. Iqbal and J. Aggarwal. Retrieval by classification of images containing large manmade objects using perceptual grouping. *Pattern Recognition*, 35:1463–1479, July 2002.

[9] P. D. Kovesi. Edges are not just steps. In *Proceedings of the Fifth Asian Conference on Computer Vision*, pages 822–827, January 2002. Melbourne.

[10] D. D. Lewis. Evaluating and optimizing autonomous text classification systems. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 246–254. ACM Press, 1995.

[11] J. Li and J. Z. Wang. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1075–1088, 2003.

[12] A. Pope and D. Lowe. Vista: A software environment for computer vision research. In *CVPR94*, pages 768–772, 1994.

[13] B. Schölkopf, O. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, and V. Vapnik. Comparing support vector machines with gaussian kernels to radial basis function classiers. *IEEE Trans. on Signal Processing*, 2(11):2758–2765, 1997.

[14] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer, N.Y, 1995.

[15] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-sensitive integrated matching for picture LIbraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9):947–963, 2001.