AN AUTOMATIC LABELING SYSTEM

Martin Zukal
Doctoral Degree Programme (1), FEEC BUT
E-mail: martin.zukal@phd.feec.vutbr.cz

Supervised by: Vít Novotný
E-mail: novotnyv@feec.vutbr.cz

Abstract: This contribution deals with the object recognition in images and assigning textual annotations (labels) to them. In order to achieve better results a combination of three frequently used approaches is chosen and implemented utilizing machine learning techniques, object ontology and incorporation of relevance feedback. The proposed system is tested on a number of object recognition tasks. The results of these tasks are included in the text as well. The system proved to be well designed but worth improving.

Keywords: content-based image retrieval, machine learning, data mining, graph theory, feature extraction

1 INTRODUCTION

The size of the image collections (personal ones as well as public ones like Flickr[1]) has grown rapidly over the last years. It is due to the development of the Internet and availability of image capturing devices [1].

The need of effective searching algorithms grows along with the growth of the number of images in the collections. There are two basic approaches how to deal with image retrieval: text-based and content-based. The former utilizes textual annotations and database management systems to retrieve the images according to the query. However, this approach suffers from two main disadvantages. Adding annotations manually can be very time-consuming and the annotations can be subjective and therefore inaccurate [1]. On the contrary, systems that are able to perform retrieval that is based on actual content of the image are referred to as the content-based image retrieval (CBIR) systems.

2 CONTENT-BASED IMAGE RETRIEVAL SYSTEMS

The existing content-based image retrieval systems process the image in a number of phases. The low-level features are extracted from the image in the initial step of the process. Many low-level feature extraction algorithms have been designed and their results have been described in a large number of articles. Features that are used very frequently are color, texture, spatial location and shape, but novel features are still needed [2]. The extracted low-level features are related to human semantics to improve the accuracy of the retrieval. The image retrieval systems often fail in relating low-level features to semantic characterization. The discrepancy between the low-level features and the richness of human semantics is referred to as the "Semantic gap" [1].

We can distinguish three major categories of techniques that are used to narrow down the semantic gap [3]:

- utilization of machine learning methods to associate low-level features with query concepts;

[1] http://www.flickr.com

- utilization of object ontology to define high-level concepts;
- utilization of relevance feedback to learn users' intention.

3 OUR APPROACH

We believe that the most accurate results can be achieved only when a combination of all three approaches is used. Our approach is based on utilization of machine learning algorithms followed by image segmentation and description of the relationships between the segments with an undirected weighted graph. Afterwards, the object ontology is used to improve the classification performed by the learning algorithm.

Machine learning techniques are used to obtain high-level semantics based on the low-level features. There are two basic types of machine learning techniques [4]: supervised learning and unsupervised learning. Supervised learning aims at predicting the value of an outcome measure (e.g. semantic category label) based on a set of input measure (i.e. the low-level features). In unsupervised learning, on the contrary, there is no outcome measure, and the goal is to describe how the input data are organized or clustered. From many existing unsupervised learning algorithms the Support Vector Machines (SVM) [5] seems to be very promising one.

The segmentation can be either complete or partial [6]. In the former an image is divided into separate homogeneous regions. The homogeneity can lie in brightness, color, texture, etc. To achieve complete segmentation of a complex scene cooperation with higher levels of processing is necessary.

Therefore we introduce a graph representation of the partially segmented image. A graph [7] is a pair $G = (V, E)$ of sets, where V represents the set of vertices (or nodes) of the graph G and elements of E are its edges. We shall assume that $V \cap E = \emptyset$. If a weight is assigned to each edge the graph is referred to as a weighted graph. In our case, the weight reflects how large the common area of the segments is. The edges can be found only between vertices representing objects that neighbor with each other.

After that the graph is related to semantics that is described with utilization of the object ontology [8]. The so-called "object ontology" is in essence a simple vocabulary of intermediate-level descriptors which provide qualitative definition of high-level concepts. By the term high-level concepts we understand abstract objects from real world like sky, lake, forest etc. With utilization of this ontology, for example lake can be described as "low, uniform, and blue region", where low refers to spatial location, uniform refers to texture and blue refers to color feature.

Finally, during the actual retrieval, the user of the system is brought in the retrieval loop to reduce the semantic gap. This is done by means of so-called relevance feedback [9]. The idea behind relevance feedback is to show the user a list of images retrieved after the initial search, ask the user to judge the results (whether each image is relevant or irrelevant), and modify the parameters of the underlying system to accommodate users' intentions. This process can be repeated and the results are refined in each iteration to provide the user with best possible results.

The whole process (feature extraction, segmentation, graph representation and object ontology) was implemented in the Rapid Miner platform which will be described in next section.

4 RAPID MINER

Rapid Miner[2] is the world-leading open-source tool for data mining. The first version has been developed at the University of Dortmund and it is available under AGPL license. Number of users all around the world reaches over hundred thousand. Rapid Miner includes hundreds of methods that

[2] http://rapid-i.com

can be used for data loading, data modeling and data visualization. It also includes an extensive set of learning methods (almost 250 different data modeling algorithms).

The design of Rapid Miner is based on concept of modular operators which define an input and an output. The operators can be placed one after another and connected together. Some operators can be placed inside other operators. The connected operators are referred to as a tree of operators. Leaves in this tree represent simple operations while inner nodes (with the degree of at least one) represent more complex or abstract steps. The XML (eXtensible Markup Language) is used as a means for description of the tree of operators.

4.1 IMAGE PROCESSING EXTENSION

Although Rapid Miner includes a lot of data mining methods it lacks the support for image processing and extraction of features from images. Our main objective was to address the absence of image processing methods and to develop an extension that will provide number of methods for advanced image processing and feature extraction from images. The extracted features can be used as an input for other (already available) operators that will classify the images in different classes or perform other data mining operations.

By now, the developed extension [10] includes over 80 operators that are divided into following groups:

- input/output operations,
- preprocessing,
- feature extraction,
- segmentation,
- visualization.

The group of preprocessing operators includes number of linear as well as nonlinear (e.g. median) filters, conversions between different color models (currently supported color models are RGB, HSV, IHLS, YUV, CIELab and CIELuv), denoising operators etc. Feature extraction operators comprise many operators related to medical image processing (e.g. Block difference of inverse probabilities – BDIP, Block variation of local correlation coefficient – BVLC) as well as operators commonly used in object detection (the so-called Haar-like features). The edge detection segmentation is an example of a simple segmentation method while Markov Random Fields (MRF) is an example of an advanced one. Operators that allow us to view the results can be found in the group of visualization operators.

5 RESULTS

We selected and implemented a number of tasks that are popular in the field of object recognition to test our system.

One of these tasks was to design and test a new interest point detector [11]. An interest point detector selects few points in the image that have unique characteristics. These points are referred to as the interest points. They are selected at different locations in the image (e.g. corners, T-junctions). Repeatability is used as a means for evaluation of an interest point detector. It denotes that detection is independent of changes in the imaging conditions like the scale (zoom), the different viewing angle, the illumination conditions etc.

The algorithm was tested on a set of images of different scene types. The size of the images was set to approximately 500x330px (the second dimension varies a little in order to keep the aspect ratio). The repeatability of the detector varied a little according to scene type. The overall results are summarized in table 1. As the table shows the interest point detector performs well for transformations in which the luminance does not vary a lot (Gaussian blur and Histogram equalization). Other transformations are prone to give non-satisfactory results. The worst repeatability score was achieved for rotation. Different rotation angles were tried, nonetheless best results were achieved for rotation of 180° although the score is very low.

Table 1: Obtained results

Algorithm	**Repeatability (percentage)**
Histogram equalization	73.3%
Gaussian blur	62.2%
High boost	60%
Erosion	17.8%
Dilation	13.3%
Rotation	2.2%

Dilation and erosion are methods in the shape analysis but they also significantly influence the appearance of the image changing the light or dark objects rapidly. In general, dilation increases the sizes of objects, filling in holes and broken areas, and connecting areas that are separated by spaces smaller than the size of the structuring element. When applied on grayscale image, dilation increases the brightness of objects by taking the neighborhood maximum when passing the structuring element over the image. Erosion, on the contrary, decreases the sizes of objects and removes small anomalies by subtracting objects with a radius smaller than the structuring element. With grayscale images, erosion reduces the brightness of bright objects on a dark background by taking the neighborhood minimum when passing the structuring element over the image [12].

The other two tasks were sky area identification in images [13] and water area identification in images [14]. The low level features were used as an input for a learning algorithm (SVM in both cases). We were very successful in the former task (the model achieved accuracy over 95% on validation data set), on the contrary, the latter task proved to be rather difficult and the results (the model achieved accuracy only 67% on validation data set) are not as good and thus will be subject to improvements.

6 FUTURE WORK

Currently, our project covers the feature extraction phase as well as the segmentation phase along with some supporting operations (preprocessing, input/output operations). Our work will continue with implementation of the graph representation and development of the vocabulary for object ontology. After that, we will design the graphical user interface that will allow us to incorporate the relevance feedback.

We plan to utilize the Tiny Images Dataset [15] which contains over 79 million images with resolution of 32x32 pixels for training our learning algorithms and testing the system results. A small subset of this huge dataset contains manual annotations so they can be used for the evaluation of the accuracy of the proposed system.

7 CONCLUSION

The obtained results described in section 5 show that our approach can be successfully used in object recognition tasks and they are strong motivation for our further work. The final solution can be

utilized as a sophisticated automatic labeling system for video sequences and images. With such a system we will easily be able to label images in large image collections. Search engine that will allow searching in movies or to automatically add annotations to scenes can be mentioned as another possible utilization.

REFERENCES

[1] Liu et al. *A survey of content-based image retrieval with high-level semantics*, Pattern Recognition 40 (1) (2007), pp. 262-282.

[2] Lew, M.S.; Sebe, N.; Djeraba, C.; Jain, R. *Content-based multimedia information retrieval: state of the art and challenges*. ACM Trans. Multimedia Comput. Commun. Appl. 2(1), 1–19 (2006)

[3] Hu Min; Yang Shuangyuan. *Overview of content-based image retrieval with high-level semantics*, Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on , vol.6, no., pp.V6-312-V6-316, 20-22 Aug. 2010

[4] J. Han; M. Kamber; Jian Pei. *Data Mining: Concepts and Techniques*, Second Edition, The Morgan Kaufmann Series in Data Management Systems.

[5] E. Chang; S. Tong. *SVMactive-support vector machine active learning for image retrieval*, Proceedings of the ACM International Multimedia Conference, October 2001, pp. 107-118.

[6] Šonka, M.; Hlaváč, V.; Boyle, R. *Image Processing, Analysis, and Machine Vision*, 3rd Edition. Toronto, Canada: Thomson Engineering, 2007, 829 s.

[7] R. Diestel. *Graph Theory*, Fourth Edition 2010, Springer-Verlag, Heidelberg, Graduate Texts in Mathematics, Volume 173, ISBN 978-3-642-14278-9, July 2010

[8] V. Mezaris; I. Kompatsiaris; M.G. Strintzis. *An ontology approach to object-based image retrieval*, Proceedings of the ICIP, vol. II, 2003, pp. 511–514.

[9] X.S. Zhu; T.S. Huang. *Relevance feedback in image retrieval: a comprehensive review*, Multimedia System 8 (6) (2003) 536–544.

[10] R. Burget; J. Karásek; Z. Smékal; V. Uher; and O. Dostál. *Rapidminer image processing extension: A platform for collaborative research*, in The 33rd International Conference on Telecommunication and Signal Processing, TSP 2010, 2010, pp. 114–118.

[11] Zukal, M.; Qui, X. *Algorithms for Interest Points Detection*. In 6th International Conference on Teleinformatics 2011 - ICT 2011, 2011. pp. 163-167.

[12] W. Landsman. The idl astronomy user's library @ONLINE. [Online]. Available: http://idlastro.gsfc.nasa.gov/

[13] R. Burget; F. Dongmei. *Identification of sky area in images according to low level features*, in Proceeding of the 6th International Conference on Teleinformatics - ICT 2011, 2011, pp. 168-173.

[14] P. Cika; F. Dongmei. *Water area identification based on image low-level features*, in Proceeding of the 6th International Conference on Teleinformatics - ICT 2011, 2011, pp. 193-196.

[15] A. Torralba; R. Fergus; W. T. Freeman. *80 million tiny images: a large data set for nonparametric object and scene recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pp. 1958–1970, 2008.