Ročník 2011 electroscope Číslo II

Utilization of Interest Point Detectors in Content Based Image Retrieval

M. Zukal [1], P. Číka[1]
[1] Department of Telecommunications, Faculty of Electrical Engineering, BUT, Brno,
Purkyňova 118, Brno
E-mail : martin.zukal@phd.feec.vutbr.cz, cika@feec.vutbr.cz

Abstract:
The paper deals with content based image retrieval in general and interest point detectors as one of possible methods used in object recognition. There are described current trends in narrowing down the so called "Semantic gap" and we propose a new solution how to achieve it. Furthermore, we evaluate three common interest point detectors with respect to the change of image brightness. Particularly, the impact of histogram equalization, brightening and darkening of the image on the repeatability of the detectors is evaluated.

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.

The system has to be able to recognize objects in the image in order to retrieve the images successfully. It has recently been shown that the interest points can be utilized for this purpose [2]. An interest point is a point in the image where the signal changes in two dimensions (for example corners, junctions, black dots in white background, etc.). There are three categories of interest point detectors in the literature, which include contour based, intensity based and parametric model based methods. Most of those methods are based on one of the basic principles: the Harris [3], the Harris-Laplace [4] and the Laplacian-of-Gaussian [5] interest point detectors.

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 [6]. 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 [7]:

- utilization of machine learning methods to associate low-level features with query concepts;
- utilization of object ontology to define high-level concepts;
- utilization of relevance feedback to learn users' intention.

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 [8]: 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

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

unsupervised learning algorithms the Support Vector Machines (SVM) [9] seems to be very promising one.

The segmentation can be either complete or partial [10]. 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 [11] 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 [12]. 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 [13]. 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.

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 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.

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 [14] includes over one hundred 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.

RESULTS

The work on content based image retrieval system still continues but there are not any meaningful results yet. We are now in the phase of testing the proposed system. We successfully implemented a number of tasks that are popular in the field of object recognition to test the system.

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

Two of the tasks were sky area identification in images [15] and water area identification in images [16]. 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.

We decided to test currently widely used interest point detectors to see whether they will be suitable for our purposes or whether we will have to design and implement our own interest point detector. Particularly, the Harris-Laplace detector [4], the Fast Hessian detector [2] and the Difference of Gaussian (DoG) detector [5] were tested. The results we obtained when testing the detectors with respect to the change of lightening conditions are described in following sections.

Evaluation of Interest Point Detectors

The repeatability rate is used to evaluate the described methods. The repeatability rate $r_i(\varepsilon)$ for image I_i is defined as

$$r_i(\varepsilon) = \frac{|R_i(\varepsilon)|}{min\,(n_1, n_i)}, \qquad (1)$$

where R_i equals to the number of point pairs (x_1, x_i) which correspond within an ε −neighborhood, n_1 and n_i are the number of points detected in common part of images I_1 and I_i. The detailed description of the repeatability rate is described in [17].

Two databases were used to evaluate the interest point detectors. The first database was The USC-SIPI Image Database volume Miscellaneous[3]. We will refer to this database as to the USC-SIPI Image Database in what follows. This collection contains thirty eight color and grayscale images with different sizes, depicting different scenes. The sizes varied between 256x256 and 512x512 pixels. Since the interest point detectors are able to work only with grayscale images the color images were converted to grayscale.

The second database was the Background category of the archive of the Caltech Computational Vision Group[4]. This database will be further referred to as the Caltech database. This collection contains four hundred and fifty one images of assorted scenes. There are fifty four images of size 378x251 pixels and three hundred and ninety seven images of size of 223x147 pixels.

[3]available at http://sipi.usc.edu/database/database.php?volume=misc

[4]available at http://www.vision.caltech.edu/html-files/archive.html

The repeatability of selected interest point detectors was investigated in three different scenarios. Firstly, we tested the influence of the decrease of light in the image on the repeatability. Secondly we tested how the increase of light influences the repeatability. Finally, the repeatability was measured after performing histogram equalization.

Darkening and brightening the image

Darkening of the image was achieved by scaling the Red, Green and Blue components of the image with a certain scale factor. To put it more simply, every value in each color component matrix was multiplied by a number (less than one in this case) and clipped to minimum value. We refer to this number as to the darkening factor. Brightening of the image was performed in a similar way. The only difference here is that the number (brightening factor) is greater than one and the resulting values are clipped to maximum value (value of 255 in the 8 bit representation).

Histogram equalization

The description of histogram equalization is a little more complicated. Let us consider a discrete grayscale image I and let n_i be the number of occurrences of gray level i. The probability that a pixel of level i occurs in the image is

$$p_I(i) = p(x = i) = \frac{n_i}{n}, \quad i \in 0, 1, \ldots, L-1 \qquad (2)$$

where L is the total number of gray levels in the image, n is the total number of pixels in the image, and $p_I(i)$ is the image's histogram for pixel value i, normalized to $[0,1]$.

The distribution function (sometimes also referred to as the cumulative distribution function -- CDF) corresponding to p_I is defined as

$$F_I(i) = \sum_{j=0}^{i} p_I(j) \qquad (3)$$

This equation also describes the image's accumulated normalized histogram.

Now, we will define a transformation of the form $y = T(x)$ which will produce a new image J, whose CDF will be linear across the range of all values

$$F_J(i) = iK \qquad (4)$$

for some constant K.

Since the CDF is an increasing function it allows us to perform such a transform. It is defined as

$$y = T(x) = F_I(x). \qquad (5)$$

Observe that the transform T maps the levels into the range $[0,1]$. To map the values into their original range, we need to apply the following transformation on the result

$$y = y \cdot (max\{x\} - min\{x\}) + min\{x\} \quad (6)$$

The tone curves of the image are changed and the details in the flat regions of the image are brought up during the histogram equalization.[18]

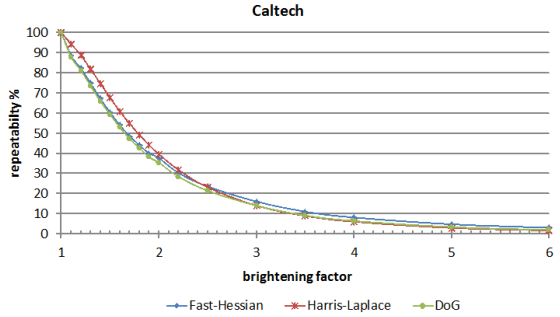

Fig. 1: Brightening - The Caltech database

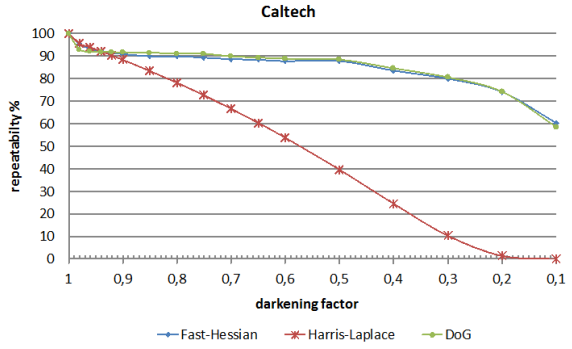

Fig. 2: Darkening - The Caltech database

We created three processes to evaluate each scenario in detail. The testing was performed on a personal computer with an Intel Core 2 CPU (Pentium R @ 2,8GHz) with 4GB of RAM. The adjustable parameters of interest point detectors were chosen so that the number of interest points detected by each detector was similar.

Seventeen values for different brightening factors were collected during the brightening test. The darkening test was performed for sixteen different values of the darkening factor. The impact of the change of the brightness of the images on repeatability is shown in figures 1 to 4. It can be seen that the Fast Hessian and DoG detectors perform very similarly while the Harris-Laplace detector slightly outperforms the first two detectors in the brightening test. On the other hand, in the darkening test the results of Harris-Laplace detector are rather poor in comparison to the other two detectors.

The results of histogram equalization test are gathered in tables 1 and 2.

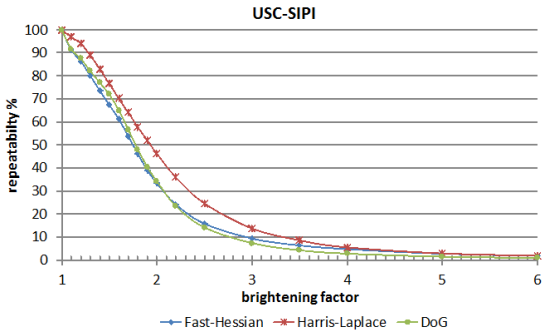

Fig. 3: Brightening - The USC-SIPI Image Database

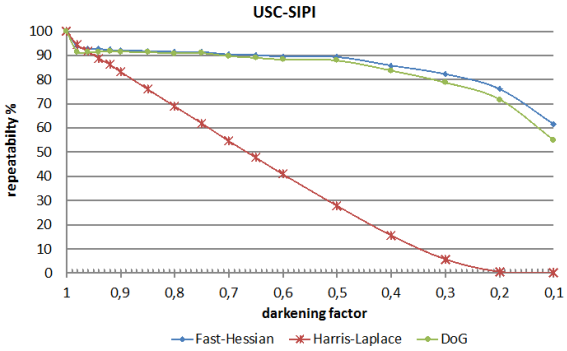

Fig. 4: Darkening - The USC-SIPI Image Database

Histogram Equalization – The Caltech Database

	Fast Hessian	***DoG***	***Harris-Laplace***
Repeatability %	65,39	56,3	71,99

Histogram Equalization – The USC-SIPI Image Database

	Fast Hessian	***DoG***	***Harris-Laplace***
Repeatability %	64,45	53,9	72,53

The difference in the behavior is caused by the fact that the Harris-Laplace detector is a corner detector while the DoG and Fast Hessian detectors respond to blob-like areas. Blobs are vague in shape but usually larger in size and more significant in terms of the change of intensity between the neighborhood of the blob and the blob itself. In other words, the corners are less and less visible during the darkening of the image while blobs remain visible in the image. This is, however, not true for brightening process. The corners as well as the blob-like areas remain visible even for quite large brightening factors.

CONCLUSION

In this article, the concept of content based image retrieval was described. Special attention was paid to the utilization of interest points which seem to

be a promising direction in the field of object recognition and consequently in the content based image retrieval itself. To decide which interest point detector from a wide range of currently available detectors will be used in the system we evaluated three of them on two different databases that contain altogether four hundred and eighty nine images. Particularly, the Harris-Laplace, the Fast Hessian and the Difference of Gaussian detectors were tested.

The interest point detectors performed very similarly. The only difference was the behavior of the Harris-Laplace detector during the darkening test. The Fast Hessian slightly outperforms the DoG detector and it is also faster in processing (according to our measurements). Therefore we will incorporate the Fast Hessian detector in our system.

We plan to utilize the Tiny Images Dataset [19] which contains over 79 million images with resolution of 32x32 pixels for further testing of the system. A small subset of this huge dataset contains manual annotations so they can be used for the evaluation of the accuracy of the image retrieval.

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] Bay H., Tuytelaars T., Van Gool, L. "Surf: Speeded up robust features," In ECCV, pages 404-417, 2006.

[3] Harris, C., Stephens, M., "A Combined Corner and Edge Detection," Proceedings of The Fourth Alvey Vision Conference, pages 147-151, 1988.

[4] Shi, F., Huang, X., Duan, Y., "Robust Harris-Laplace Detector by Scale Multiplication," Advances in Visual Computing in Lecture Notes in Computer Science, pages 265-274, 2009. Springer Berlin / Heidelberg.

[5] Lowe, D., "Distinctive Image Features from Scale-Invariant Keypoints," Journal of Computer Vision, 2(60), 91-110, 2004.

[6] 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)

[7] Hu M., Yang S., "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

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

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

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

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

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

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

[14] Burget, R., Karásek, J., Smékal, Z., Uher, V., and Dostál, O., "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.

[15] Burget, R., Fu, D.M., "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.

[16] Cika, P., Fu, D.M., "Water area identification based on image low-level features," in Proceeding of the 6th International Conference on Teleinformatics - ICT 2011, 2011, pp. 193-196.

[17] Schmid, C., Mohr, R., Bauckhage, C., "Evaluation of Interest Point Detectors," Int. J. Comput. Vision, 37:151--172, 2000.

[18] Acharya, T. and Ray. A.K., *Image Processing: Principles and Applications*, Wiley-Interscience 2005, ISBN 0-471-71998-6

[19] Torralba, A., Fergus, R., Freeman, W. T., "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.