Face-based Image Retrieval – One Step Toward Object-based Image Retrieval *

Thomas Deselaers, David Rybach, Philippe Dreuw
Daniel Keysers[2], and Hermann Ney
Lehrstuhl für Informatik VI, Computer Science Department
RWTH Aachen University, D-52056 Aachen, Germany
Tel.: +49 241 8021613, Fax.: +49 241 8022219
e-mail: deselaers@cs.rwth-aachen.de
[2]German Research Center for Artificial Intelligence (DFKI), www.iupr.org

Abstract

In this paper we propose a method to retrieve images based on the persons shown. The method aims at retrieving from images showing groups of people those in which the same persons are depicted as in the query image. It is experimentally shown that this aim is achieved for rather simple tasks and that improvements over baseline methods are possible for harder tasks.

1 Introduction

Content-based image retrieval (CBIR) is a hot topic in research, but only few articles have been written about what people expect from an image retrieval system. Those articles that are concerned with the users' needs focus on journalists [12] and medical doctors [9].

The most common request from journalists to picture archives are queries for certain objects. Among these, the most frequently issued queries ask for certain persons. Unfortunately, no current CBIR system can handle general photographs of persons.

In this paper we present a method that combines techniques from face detection and recognition with techniques from content-based image retrieval to allow the retrieval of images showing certain persons. A first, though very limited, approach to achieve this goal has been implemented in the PhotoBook system [15] which allowed for retrieving images of persons when the face covered a large part of the image and was taken under normalized conditions. In [2] an approach to the same problem is proposed. A more complex approach to retrieving portrait images robust with respect to facial expression, glasses, hats, and facial hair is presented in [13].

*This work was partially funded by the DFG (Deutsche Forschungsgemeinschaft) under contract NE-572/6.

A component-based descriptor for face images using LDA transformation is presented in [11]. In [1] portrait images taken from the web are indexed and performance measures for comparing such systems are proposed. All of these approaches have in common that there is only one face in the image. Furthermore it is assumed that the face to be recognized forms a significant part of the image, which is not a suitable assumption for general image retrieval tasks. In contrast to these approaches we propose a method that is able to deal with images displaying several persons and in which the faces do not necessarily form large parts of the image.

To achieve this goal, we extend our CBIR system FIRE[1] [4] by a new *face feature*. These features are extracted from images using the Viola & Jones face detection method [18] and Eigenface representations [17]. This face feature can then be combined with all image representation features already available in FIRE.

Considering faces as objects in the general sense, this work also opens up new vistas to retrieving images based on the objects they contain. The "only" necessity therefore is to have a reasonably well working object detection and classification algorithm that is combined with the image retrieval engine. In the field of object classification and detection recently many approaches were presented [14, 8, 10, 7, 5].

2 The Image Retrieval Framework

The FIRE system [4] follows the approach of representing images by feature vectors that are compared using certain distance measures. Another approach is inspired by textual information retrieval and images are represented by a set of binary features. This approach is e.g. followed by the GIFT system [16].

Given a set of positive example images Q^+ and a (possibly empty) set of negative example images Q^- a score $S(Q^+, Q^-, X)$ is calculated for each image X from the database $\mathcal{B}$:

$$S(Q^+, Q^-, X) = \sum_{q \in Q^+} S(q, X) + \sum_{q \in Q^-} (1 - S(q, X)). \tag{1}$$

where $S(q, X)$ is the score of database image X with respect to query q and is calculated as $S(q, X) = e^{-\gamma D(q,X)}$ with $\gamma = 1.0$. $D(q, X)$ is a weighted sum of distances calculated as

$$D(q, X) := \sum_{m=1}^{M} w_m \cdot d_m(q_m, X_m). \tag{2}$$

Here, q_m and X_m are the mth feature of the query image q and the database image X, respectively. d_m is the corresponding distance measure and w_m is a weighting coefficient. For each d_m, $\sum_{X \in \mathcal{B}} d_m(Q_m, X_m) = 1$ is enforced by re-normalization.

Given a query (Q^+, Q^-), the images are ranked according to descending score and the K images X with highest scores $S(Q^+, Q^-, X)$ are returned by the retrieval system.

Due to the lack of suitable training data, weights w_m were chosen heuristically based on experiences from earlier experiments with other data.

[1] available online at http://www-i6.informatik.rwth-aachen.de/~deselaers/fire.html

3 Face Detection and Representation

To allow for investigating the effect of face detection and representation methods, we tested two methods: detection and representation using Eigenfaces [17] and detection using the Viola & Jones method [18] and representing the face as a size normalized image patch. Both methods are data driven methods and are explained briefly in the following. A general overview on methods for face detection and representation can be found in [19].

3.1 Eigenfaces

Turk and Pentland applied principal component analysis to face recognition and detection [17]. Principal component analysis is performed on a training set of face images to generate the Eigen-vectors (here called Eigenfaces) which span a subspace (called the face space) of the image space. To detect whether an image shows a face or not, it is projected into the subspace and back-projected using e.g. only the 20 first face space components. Then, the distance between the original image and the back-projection can be calculated. Due to the nature of the Eigenfaces, face-like images can be reconstructed well, whereas non-faces are reconstructed very poorly and thus the distance between the original image and the back-projection is high in this case. Thus, this distance can be seen as a measure of *faceness*. If faceness is calculated for every position in the image, a face map can be generated and a face can be detected by searching for local minima. An advantage of this method is that it directly offers a compact and generalizing method of representing faces.

3.2 Viola & Jones Method

Viola and Jones present a new and radically faster approach to face detection based on the AdaBoost algorithm from machine learning [18]. Boosting is a method of combining several weak classifiers to generate a strong classifier. AdaBoost is a well known algorithm to generate strong classifiers from weak classifiers, while providing statistical bounds on the training and generalization error of the algorithm. The weak classifiers in the Viola & Jones algorithm are based on features of three kinds. A two-rectangle feature is the difference between the sum of the values of two adjacent rectangular windows. A three-rectangle feature considers three adjacent rectangles and computes the difference between the sum of the pixels in the extreme rectangles and the sum of the pixels in the center rectangle. A four-rectangle feature considers a 2×2 set of rectangles and computes the difference between the sum of the pixels in the rectangles that constitute the main and off diagonals. For a 24×14 sub-window there could be more than 180,000 such features.

3.3 Faces in Image Retrieval

To be able to use the faces in our image retrieval framework, we have to define a distance measure for two images X and Y in which faces $X_1 \dots X_F$ and $Y_1 \dots Y_V$ have been detected. One can think about various ways of matching, e.g. taking into account the positions of the

faces. For the experiments presented here, we decided to use the simplest possible matching. That is, we calculate the Euclidean distances $d(X_i, Y_j)$ between all pairs (X_i, Y_j) of faces of two images X and Y. The smallest of these distances $d(X_i, Y_j)$ is then used as the distance between the images X and Y. This can be interpreted as retrieving images that contain one of the persons in the query image, and this person's face should be as similar as possible to her face in the query image.

4 Databases

We present experiments on two different databases:

a) Bio-ID database: a database showing one person per image which can be said to be recorded under controlled conditions.

b) RWTH-i6 Groups of People Database: a newly created database which has been collected using Google image search. It is obvious that this is a much harder task.

4.1 Bio-ID

The dataset consists of 1521 gray level images with a resolution of 384×286 pixel. Each image shows the frontal view of a face of one out of 23 different test persons. The database is available online[2]. This task is very simple and the results appear to be very good. Unfortunately, the data are not properly labelled, thus a quantitative evaluation is not easily possible. Example images are given in Figure 1.

4.2 RWTH-i6 Groups of People Database

Due to the lack of a database containing images of groups of persons, images of single persons, and images without persons at all, we decided to create our own database. To do so we queried Google image search[3] using names of politicians, musicians, and music bands (e.g. Gerhard Schroeder, Britney Spears, Depeche Mode) and retrieved 60 images for each of these keywords. In total we used 38 search terms. Then we deleted all images that were not relevant to the search term, as e.g. images showing dogs, comics, or images showing other people. This lead to a database containing 868 images with 38 classes. Some example images along with their class label are given in Figure 1.

5 Experimental Results

In the following we present some exemplary results of using our image retrieval system FIRE to retrieve images from a database containing mainly images of people. We compare a setting

[2] http://www.humanscan.de/support/downloads/facedb.php
[3] http://images.google.com

Figure 1: Example images from the Bio-ID database and from our newly created “Groups of People” database.

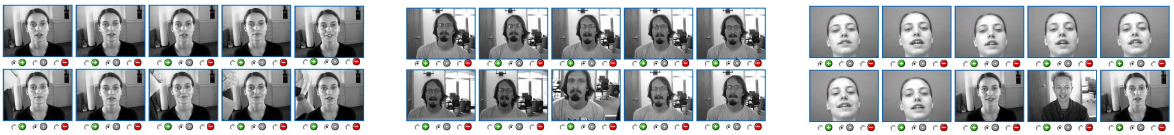

Figure 2: Example results of queries in the Bio-ID database where each top-left face image was the query image.

that is known to perform well on general photographs with the same setting augmented by the described face features.

5.1 Bio-ID database.

As it is impossible to give quantitative results for the Bio-ID database we show just three queries, although we are aware that this is not the best practice.The queries are exemplary and most queries give similar results. Examples are shown in Figure 2 where the top-left image is the query image in each of the cases. They clearly show that among the first 10 images retrieved, most images are relevant. As the results here are very good, we proceed to the harder problem with our newly created database. For the experiments here we used the implementation of the Viola & Jones face detector from OpenCV[4]. Detected faces are extracted, scaled to 16×16 pixels and used as features.

[4] http://www.intel.com/technology/computing/opencv/index.htm

Table 1: Results for the experiments performed

Setup	ER	MAP
color histograms & global texture feature	78.2	6.55
inv. feat. histograms & Tamura texture feature	78.3	6.95
16x16 faces, normalized	73.0	7.35
19x19 faces (center part), normalized	73.5	7.53
19x19 faces (center part), normalized, 10 Eigenface coeff	71.0	8.22

5.2 RWTH Groups of People database.

To evaluate the quality of the proposed method we compare it to two different baseline systems:

1. A combination of color histograms and a cooccurrence-based texture feature. A very similar feature combination has been used in QBIC [6].

2. A combination of an invariant feature histogram accounting for color and Tamura texture histograms. This method has been shown to perform well on general photographs [3].

Then we performed a series of experiments with the proposed method where the faces were extracted and represented in different ways.

All experiments were performed in a leaving-one-out manner. That is, each image is used as query image and the remaining images are used as database, images from the same class are considered to be relevant. Results are given as error rates (as due to the limited number of classes this can be considered to be a classification problem and as MAP (mean average precision) which better reflects the retrieval precision over the complete ranking of database images.

Results for different setups are given in Table 1. It can be seen that the results are far from satisfactory. Nonetheless using face features, a considerable improvement can be achieved.

The high error rates are due to problems with the recognition of similar faces. The detection of the faces works reasonable well. Figure 3 shows some examples of detected faces. So improvements are necessary in the area of comparing faces and recognizing whether two faces are similar or not. In the domain of face verification these problems are partly solved already and we are currently incorporating this knowledge into our system.

6 Conclusion

We have presented an approach to retrieving general images based on faces depicted. The approach works reasonably well when faces contained in the database are sufficiently similar to the query images which is the case for the Bio-ID database. For the newly created database showing groups of people, results are not yet satisfactory but a considerable improvement over the baseline system for general images could be obtained. In combination with knowledge from face recognition and face verification it is very probable that strong improvements are possible.

Figure 3: Examples of face detection results.

References

[1] Yuksel Alp Aslandogan and Clement T. Yu. Evaluating strategies and systems for content based indexing of person images on the web. In *ACM International Conference on Multimedia*, pages 313–321, Los Angeles, CA, USA, October/November 2000.

[2] Nozha Boujemaa and Chahab Naster. Content-based image retrieval at the imedia group of inria. Technical report, INRIA, Montbonnot, France, 2000.

[3] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Features for image retrieval – a quantitative comparison. In *DAGM 2004, Pattern Recognition, 26th DAGM Symposium*, number 3175 in LNCS, pages 228–236, Tübingen, Germany, September 2004.

[4] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Fire – flexible image retrieval engine: Imageclef 2004 evaluation. In *Multilingual Information Access for Text, Speech and Images. Proceedings of the 5th Workshop of the Cross-Language Evaluation Forum. CLEF 2004*, volume 3491 of *LNCS*, pages 688–698, Bath, UK, September 2004. Springer.

[5] Thomas Deselaers, Daniel Keysers, and Hermann Ney. Discriminative training for object recognition using image patches. In *CVPR 05*, volume 2, pages 157–162, San Diego, CA, jun 2005.

[6] Christos Faloutsos, Ron Barber, Myron Flickner, Jim Hafner, Wayne Niblack, Dragutin Petkovic, and William Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 3(3/4):231–262, July 1994.

[7] Li Fei Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 524–531, San Diego, CA, USA, June 2005. IEEE.

[8] Rob Fergus, Pietro Perona, and Andrew Zisserman. A sparse object category model for efficient learning and exhaustive recognition. In Cordelia Schmid, Stefano Soatto, and

Carlo Tomasi, editors, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 380–389, San Diego, CA, USA, June 2005. IEEE.

[9] William Hersh, Jeffery Jensen, Henning Müller, Patrick Ruch, and Paul Gorman. A qualitative task analysis of biomedical image use and retrieval. In *Proc. MUSCLE/ImageCLEF Workshop 05*, Vienna, Austria, September 2005.

[10] Alex Holub and Pietro Perona. A disriminative framework for modelling object classes. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 663–670, San Diego, CA, USA, June 2005. IEEE.

[11] Tai-Kyun Kim, Hyunwoo Kim, Wonjun Hwang, Seok Cheol Kee, and Jong Ha Lee. Component-based lda face descriptor for image retrieval. In *British Machine Vision Conference*, Cardiff, UK, September 2002.

[12] M. Markkula and E. Sormunen. Searching for photos - journalists' practices in pictorial IR. In *Electronic Workshops in Computing – Challenge of Image Retrieval*, pages 1–13, Newcastle, UK, February 1998.

[13] Aleix Martinez. Face image retrieval using hmms. In *IEEE Workshop on Content-Based Access of Image and Video Libraries*, pages 35–39, Fort Collins, CO, USA, jun 99.

[14] Raphaël Marée, Pierre Geurts, Justus Piater, and Louis Wehenkel. Random subwindows for robust image classification. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 34–40, San Diego, CA, USA, June 2005. IEEE.

[15] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 1995.

[16] David McG. Squire, Wolfgang Müller, Henning Müller, and Jilali Raki. Content-based query of image databases, inspirations from text retrieval: Inverted files, frequency-based weights and relevance feedback. In *Scandinavian Conference on Image Analysis*, pages 143–149, Kangerlussuaq, Greenland, June 1999.

[17] Matthew Turk and Alex Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, jan 1991.

[18] Paul Viola and Michael Jones. Robust real-time object detection. In *Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling*, pages 1–25, Vancouver, Canada, July 2001.

[19] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, Jan 2002.