International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 6 (2015) pp. 14371-14377

http://www.ripublication.com

A Real-Time Automated System For Scalable Face Image Retrieval Using Classifiers

R.Valarmathi[1] and S.Saravanan[2]
[1]Assistant Professor, Department of Electronics and Communication Engineering Jayam College of Engineering and Technology, Dharmapuri, Tamil Nadu.
[2]Professor, Department of EEE, Muthayammal Engineering college, Rasipuram, Tamil Nadu.
[1]manishmouriya2010@gmail.com [2] saravanan.nivi@gmail.com

Abstract

Image retrieval requires a system to find information relevant to a query which represents images containing faces of the same person appearing in the query image. In this paper, we aim to build a scalable face image retrieval system. For this purpose, we develop a new scalable face representation using Gray level co-occurrence matrix (GLCM) features at different orientation (45,60,90 degrees), gray level and moment invariant features, orientation histogram features and Law's texture features.

The extracted features are trained and classified by feed forward back propagation neural network and Support vector machine (SVM) classifier to rank the candidate images. The performance of the designed face image retrieval system will be analyzed in terms of Accuracy and retrieval rate. The performance of the proposed retrieval system will be compared with existing system.

Introduction

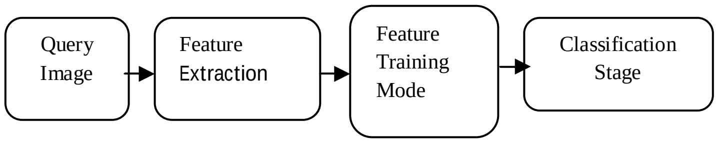
Given a face image as a query, our goal is to retrieve images containing faces of the same person appearing in the query image, from a web-scale image database containing tens of millions face images. In this paper, we assume face images are frontal with up to about 20 degrees of pose changes, such that the five face components (e.g., eyes, nose, and mouth) are visible in a given face image. Figure 1 shows some example online celebrity face images with various poses, expressions, and illumination. Such a face retrieval system has many applications, including name-based face image search, face tagging in images and videos, copyright enforcement, etc. To the best of our knowledge, little work aims at web-scale face image retrieval.

A straight-forward approach is to use the bag-of-visual words representation that has been used in state-of-the art scalable image retrieval systems [1]. However, the performance of such a system degrades significantly when applying on face images. There are two major reasons. On one hand, the visual word vocabulary, learned from local SIFT-like features detected from the face images, has difficulty achieving both high discriminative power (to differentiate different persons) and invariance (to tolerate the variations of the same person). Secondly, existing systems ignore strong, face-specific geometric constraints among different visual words in a face image. Recent works on face recognition have proposed various discriminative facial features [2], [3]. However, these features are typically high dimensional and global, thus not suitable for quantization and inverted indexing. In other words, using such global features in a retrieval system requires essentially a linear scan of the whole database in order to process a query, which is prohibitive for a web-scale image database.

Literature Survey

The State-of-the-art large scale image retrieval systems have relied on bag-of-words image representation and textual indexing and retrieval schemes for scalability. In these systems, feature detectors first detect distinctive and repeatable points or regions such as DoG (Difference of Gaussian) [4], MSER (Maximally Stable Extremal Region) [13] in the image, from which discriminative feature descriptors [5], [7] are then computed. These descriptors are subsequently quantized into visual words with a visual vocabulary [8], which is trained by the unsupervised clustering algorithms. To further improve the scalability, Jegou aggregates partial information of the standard bag-of-features vector to build the “miniBOF” (mini- Bag-of-Features) vector [9], which is more compact in the index. On the other hand, to improve the precision, some compact information can be embedded [8] for each visual word in the index, which compensates the information loss in the quantization. However, the performance of these traditional image retrieval systems degrades significantly when applied to face images. In recent years, many effective features have been proposed for face recognition. For example, LBP (Local Binary Pattern) [2] feature, variations of LBP [9],[10] and V1-like feature are designed to capture the micropatterns of the face. Besides these “low level” features mentioned above, Kumar *et al* [11] incorporates the traits information with the attribute and simile classifiers. Efforts are also made to tackle the face alignment and matching problem in face recognition. In [3], Wright proposes a Rptree (Random Projection tree) based approach to implicitly encode the geometric information into the feature. It is non-trivial to make these global feature based methods scalable. One might consider using *k*-d tree [4] or LSH (Locality Sensitive Hashing) [5], [10] to avoid scanning every image. But we have found these approximated nearest neighbor search methods do not scale or work well with high dimensional global face features.

Proposed System Model
The proposed system model consists of feature extraction module, feature selection module and the selected features are trained and classified by feed forward back propagation neural network and SVM classifier.



Feature Extraction
Features, characteristics of the objects of interest, if selected carefully are representative of the maximum relevant information that the image has to offer for a complete characterization a lesion. Feature extraction figure methodologies analyze objects and images to extract the most prominent features that are representative of the various classes of objects. Features are used as inputs to classifiers that assign them to the class that they represent.

Gray level features:
A set of gray-level-based descriptors taking this information into account were derived from homogenized images considering only a small pixel region centered on the described pixel. These features are

$$F_1(x,y)= I_H\,(x,y) - \min\{I_H\,(s,t)\} \tag{1}$$

$$F_2(x,y)= \max\{I_H\,(s,t)\} - I_H\,(x,y) \tag{2}$$

$$F_3(x,y)= I_H\,(x,y) - \mathrm{mean}\{I_H\,(s,t)\} \tag{3}$$

$$F_4(x,y)= \mathrm{std}\{I_H\,(s,t)\} \tag{4}$$

$$F_5(x,y)= I_H\,(x,y) \tag{5}$$

Moment Invariants-Based Features
The facial images are known to be piecewise linear and can be approximated by many connected line segments. For detecting these quasi-linear shapes, which are not all equally wide and may be oriented at any angle, shape descriptors invariant to translation, rotation and scale change may play an important role. Within this context, moment invariants proposed by Hu [6],[12] provide an attractive solution and are included in the feature vector. In this paper, they are computed as follows.

$$f_6(x,y)=|\log(\Phi_1)| \tag{6}$$

$$f_7(x,y)=|\log(\Phi_2)| \tag{7}$$

$$\Phi_1=\eta_{20}+\eta_{02} \tag{8}$$

$$\Phi_2=(\eta_{20}+\eta_{02})^2+4\,\eta^2_{11} \tag{10}$$

$$\eta_{pq} = \mu_{pq} / (\mu_{00})^{\gamma} \tag{11}$$

$$\gamma = \frac{p+q}{2} + 1; \tag{12}$$

$$\mu_{pq} = \sum_i \sum_j (i - i')^p (j - j')^q I_{VE}(i,j) \tag{13}$$

$$i' = m_{10} / m_{00} \tag{14}$$

$$j' = m_{01} / m_{00} \tag{15}$$

$$m_{pq} = \sum_i \sum_j (i)^{\wedge}2\ (j)^2\ I_{VE}(i,j); p,q=0,1,2\ldots.$$

Laws' Texture Energy Measures:

Laws' texture energy measures are based on convolution kernels that emphasize specific structural patterns, and could be used to generate useful features related to the intersecting tissue structures, speculations, and node-like patterns of architectural distortion. Laws [4] used several 1-D and 2-D convolution kernels to classify each pixel in an image based on measures of local "texture energy."

Table 1:

Feature Number	Feature name	Feature values
1	autocd	44.15
2	Contrd	1.89
3	Corrpd	0.1592
4	Cpromd	37.45
5	Cshad1	4.267
6	dissid	0.8876

L3 = (1, 2, 1), E3 = (-1, 0, 1), S3 = (-1, 2,-1).

Convolving these features

L5 = (1, 4, 6, 4, 1) = L3 * L3 …… (Level).

S5 = (-1, 0, 2, 0,-1) = L3 * S3..... (Spot).

R5 = (1,-4, 6,-4, 1) = S3 * S3 …..(Ripple).

E5 = (-1,-2, 0, 2, 1) = L3 * E3 …..(Edge).

W5 = (-1, 2, 0,-2, 1) = E3 * S3 …… (Wave).

GLCM features

It is a statistical method that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (*I*, *J*) in the resultant GLCM is simply the sum of the number of times that the pixel with value *I* occurred in the specified spatial relationship to a pixel with value *J* in the input image.

The Following GLCM features were extracted in our research work:

Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity Energy, Entropy, Homogeneity, Maximum probability , Sum of squares, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation, information measure of correlation, Inverse difference normalized.

Intensity Histogram Features

Intensity Histogram analysis has been extensively researched in the initial stages of development of this algorithm. Prior studies have yielded the intensity histogram features like mean, variance, entropy etc. These are summarized in Table 2. Mean values characterize individual calcifications; Standard Deviations (SD) characterize the cluster.

Table 2: Histogram features

Feature Number Assigned	Feature
1	Mean
2	Variance
3	Skewness
4	Kurtosis
5	Entropy
6	Energy

Classifiers

SVM Classifier Design

Initially designed for binary classification after supervised learning, support vector machines (SVMs) are also used for multiclass problems through one against all and one against one strategy [22]. Nonlinear class separation in low dimension space may result in smart separation in higher dimension space, using a suitable kernel function. The key point of the SVM classifier design remains the choice of the kernel function, as this depends on the image database and input descriptors [11] since no universal kernel will fit all applications. The SVM classifier selected here is a soft-margin algorithm (so-called C-SVM) available online. It has been tested by computing ROC curves for several classical kernels: linear, polynomial, radial basic function (RBF) and perceptron (Fig. 3). After the selection of a particular kernel, its parameters must be tuned. In the case of the perceptron kernel finally selected, there is only one free parameter; it controls the penalty of the classification error and has been adjusted by a line search technique.

Neural Network Classifier Design

A multilayer feed forward network, consisting of an input layer, three hidden layers and an output layer, is adopted in this paper. The input layer is composed by a number of neurons equal to the dimension of the feature vector (seven neurons). Regarding the hidden layers, several topologies with different numbers of neurons were tested. A number of three hidden layers, each containing 15 neurons, provided optimal NN configuration. The output layer contains a single neuron and is attached, as the remainder units, to a nonlinear logistic sigmoid activation function, so its output ranges between 0 and 1. Table 3 summarizes the values for those features.

Table 3: Classifier Performance

Performance Metric	SVM Classifier	NN Classifier
Accuracy Rate	0.671	0.745
Retrieval Rate (in sec)	0.16	0.13

References

[1]. Zhimin Cao, Qi Yin, Jian Sun, and Xiaoou Tang. Face recognition with learning-based descriptor and pose-adaptive matching. In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, 2010.

[2]. Jaime G. Carbonell, Yiming Yang, Robert E. Frederking, Ralf D. Brown, Yibing Geng, and Danny Lee. Translingual information retrieval: A comparative evaluation. In IN PROCEEDINGS OF THE FIFTEENTH INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, pages 708–714, 1997.

[3]. Jian Chen, Rui Ma, and Zhong Su. Weighting visual features with pseudo relevance feedback for cbir. In Conference on Image And Video Retrieval, 2010.

[4]. J. Friedman, J. Bentley, and R. Finkel. An algorithm for finding best matches in logarithmic expected time. ACM Trans. Math. Software, 3:209226, 1977.

[5]. A. Gionis, P. Indyk, and R.Motwani. Similarity search in high dimensions via hashing. In International Conference on Very Large Data Bases, 1999.

[6]. G. Hua and A. Akbarzadeh. A robust elastic and partial matching metric for face recognition. In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, 2009.

[7]. Gary B. Huang, Marwan Mattar, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In European Conference on Computer Vision, 2008.

[8]. Herve Jegou, Matthijs Douze, and Cordelia Schmid. Hamming embedding and weak geometric consistency for large scale image search. In European Conference on Computer Vision, 2008.

[9]. Herve Jegou, Matthijs Douze, and Cordelia Schmid. Packing bagof-features. In Proc. IEEE International Conference on Computer Vision, 2009.

[10]. Brian Kulis and Kristen Grauman. Kernelized locality-sensitive hashing for scalable image search. In Proc. IEEE International Conference on Computer Vision, 2009.

[11]. Neeraj Kumar, Alexander C.Berg, Peter N.Belhumeur, and Shree K.Nayar. Attribute and simile classifiers for face verification. In Proc. IEEE International Conference on Computer Vision, 2009.

[12]. Ping-Han Lee, Gee-Sern Hsu, and Yi-Ping Hung. Face verification and identification using facial trait code. In Proc. IEEE International Conference on Computer Vision and Pattern Recognition, 2009.

[13]. Z. Lei, S. Li, R. Chu, and X. Zhu. Face recognition with local gabor textons. Lecture Notes in Computer Science, 4642:49, 2007.

[14]. Lin Liang, Rong Xiao, Fang Wen, and Jian Sun. Face alignment via component-based discriminative search. In European Conference on Computer Vision, 2008.