

NCTU_DBLAB@ImageCLEF 2005: Automatic annotation task

Pei-Cheng Cheng¹, Been-Chian Chien², Hao-Ren Ke³, and Wei-Pang Yang^{1,4}

¹Department of Computer & Information Science, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
{cpc, wpyang}@cis.nctu.edu.tw

²Department of Computer Science and Information Engineering,
National University of Tainan
33, Sec. 2, Su-Lin Street, Tainan, 700, Taiwan, R.O.C.
bcchien@mail.nutn.edu.tw

³Institute of Information Management and University Library, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu, TAIWAN 30050, R.O.C.
claven@lib.nctu.edu.tw

⁴Department of Information Management, National Dong Hwa University
1, Sec. 2, Da Hsueh Rd., Shou-Feng, Hualien, Taiwan, R.O.C.
wpyang@mail.ndhu.edu.tw

Abstract

In this paper, we use Support Vector Machine (SVM) to learn image feature characteristics for assisting the task of image classification. The ImageCLEF 2005 evaluation offers a superior test bed for medical image content retrieval. Several image visual features (including histogram, spatial layout, coherence moment and gabor features) have been employed in this paper to categorize the 1,000 test images into 57 classes. Based on the SVM model, we can examine which image feature is more promising in medical image retrieval.

The result shows that the spatial relationship of pixels is a very important feature in medical image data, because medical image data always have similar anatomic regions (lung, liver, head, and so on); therefore image features emphasizing spatial relationship have better result than others.

ACM Categories and Subject Descriptors: Pattern Recognition; Computer Vision;

Free Keywords: Medical Image Classification; Support Vector Machine;

1. Introduction

Image retrieval techniques are essentially required due to the enormous amount of digital images produced day by day. Content-based image retrieval using primitive image features is promising in retrieving visually similar images. The QBIC system [Flickner95] is one of the well-known content-based systems, and other famous systems are Blob World [Carson99][Belongie98], VIPER/GIFT[Squire 99], and SIMPLcity[Wang 01]. Various image features and similarity metrics have already been proposed for general images; however, before the ImageCLEF forum, a prime problem is the lack of benchmarks to evaluate which features are suitable for a specific application..

The ImageCLEF forum offers an image test bed to compare and evaluate different visual features and distance metrics. In the automatic annotation task, a database of 9,000 fully classified radiographs in 57 classes taken randomly from medical routine is made available and can be used to train a classification system. One thousand radiographs whose classification labels are not available to the participants have to be classified. The aim is to find out how well current techniques can identify image modality, body orientation, body region, and biological system examined based on the images.

In this paper, we use the Support Vector Machine (SVM) to learn image feature characteristics. Based on the SVM model, several image features that consider, from the viewpoint of human, the invariance in image rotation, shift and illumination are employed in our system. Using these image features, the support vector machine plays as a classifier to categorize the 1,000 test images into 57 classes.

The experiment result shows that in a medical image application, an image feature with strong spatial descriptor is more significant for representing an image. The rest of this paper is organized as follows. In Section 2, the employed image features are described. Section 3 illustrates the SVM model that is used to classify the training data. In Section 4, the submissions are discussed. Finally, Section 5 provides concluding remarks and future directions for medical image retrieval.

2. Image features

This section describes the features used in this paper for the ImageCLEF 2005 evaluation. In an image retrieval system, image features are extracted from pixels. The extracted features are then used for similarity comparison. For fast response, image features must be concise, and for precision, image features must contain meaningful information to represent the image itself. Image features will directly affect the retrieval result. In this paper we examine several image features to understand which features will have good performance in medical image application.

When designing the image features, to emphasize the contrast of an image and handle images with little illuminative influence, we normalize the value of a pixel before quantization. In [Cheng 04] we proposed a relative normalization method. First, we cluster a whole image into four clusters by the K-means clustering method [Han01]. We sort the four clusters ascendantly according to their mean values. We shift the mean of the first cluster to the value 50 and the fourth cluster to 200; then, each pixel in a cluster is multiplied by a relative weight to normalize. Let m_{c1} be the mean value of cluster 1 and m_{c4} is the mean value of cluster 4. The normalization formula of pixel $p(x,y)$ is defined in Eq. (1).

$$p(x,y)_{normal} = (p(x,y) - (m_{c1} - 50)) \times \frac{200}{(m_{c4} - m_{c1})} . \quad (1)$$

After normalization, we scale an image into common 128*128 pixels and extract image features.

2.1 Facade scale image feature

The pixel values of an image are a straight-forward feature. For computational efficiency, images are always scaled to a common small size and compared using the Euclidean distance. [Keysers04] has shown that the Facade image feature performs very well in optical character recognition and medical image retrieval. In this paper we scale down an image into 8*8 pixels to form a 64 feature vectors.

2.2 Fuzzy Histogram layout

Histogram [Swain91] is a prime image feature for image retrieval. Histogram is invariant in image rotation. It is easy to implement and has good results in color image indexing. Because a radiotherapy medical image only consists of gray-level pixels, spatial relationship becomes very important. Medical images always contain particular anatomic regions (lung, liver, head, and so on); therefore, similar images have similar spatial structures. We divide an image into nine sections and calculate their histogram respectively. After normalization, gray values are quantized into 16 levels for computational efficiency.

In the fuzzy histogram layout, a gray value may be quantized into several bins to improve the similarity between adjacent bins. We set an interval range δ to extend the similarity of each gray value. The fuzzy histogram layout estimates the probability of each gray level that appears in a particular area. The probability equation is defined in Eq. (2), where δ is set to 10, p_j is a pixel of a given image, and m is the total number of pixels in this image. In our implementation, we use a total of 144 bins for the fuzzy histogram layout.

$$h_{c_i}(I) = \frac{\sum_{j=1}^m \frac{[p_j - \frac{\delta}{2}, p_j + \frac{\delta}{2}] \cap c_i}{\delta}}{m} . \quad (2)$$

2.3 Coherence Moment

One of the problems to devise an image representation is the semantic gap. The state-of-the-art technology still cannot reliably identify objects. The coherence moment feature attempts to describe the features from the human's viewpoint in order to reduce the semantic gap.

We cluster an image into four classes by the K-means algorithm. After clustering an image into four classes, we calculate the number of pixels (COH_k), mean value of the gray values (COH_μ) and standard variance of the gray values (COH_σ) in each class. For each class, we group connected pixels in the eight directions as an object. If an object is bigger than 5% of the whole image, we denote it as a big object; otherwise it is a small object. We count how many big objects (COH_o) and small objects (COH_v) are in each class, and use COH_o and COH_v as parts of image features.

Since we intend to know the reciprocal effects among pixels, so we use the smoothing method on the image. If

the spatial distribution of the pixels in two images is similar, they will also be similar after smoothing. If their spatial distributions are quite different, they may have a different result after smoothing. After smoothing, we cluster an image into four classes and calculate the number of big objects (COH_r) and small objects (COH_o). Each pixel will be influenced by its neighboring pixels. Two close objects of the same class may be merged into one object. Then, we can analyze the variation between the two images before and after smoothing. The coherence moment of each class forms a seven-feature vector, (COH_k, COH_u, COH_p, COH_o, COH_v, COH_r, COH_o). The coherence moment of an image is a 56-feature vector that combines the coherence moments of the four classes.

2.4 Gray Correlogram.

The contour of a medical image contains rich information. In this task we are going to find similar medical images. A broken bone in the contour may be different from a healthy one. Thus we choose a representation that can estimate the partial similarity of two images and can be used to calculate their global similarity.

We analyze the image pixels by our modified correlogram algorithm. The definition of the correlogram [18,19] is in Eq. (3). Let D denote a set of fixed distances $\{d_1, d_2, d_3, \dots, d_n\}$. The correlogram of an image I is defined as the probability of a color pair (c_i, c_j) at a distance d .

$$\gamma_{c_i, c_j}^d(I) = \Pr_{p_1 \in c_i, p_2 \in I} \{p_2 \in c_j \mid |p_1 - p_2| = d\} \quad . \quad (3)$$

For computational efficiency, the autocorrelogram is defined in Eq. (4)

$$\lambda_{c_i}^d(I) = \Pr_{p_1 \in c_i, p_2 \in I} \{p_2 \in c_i \mid |p_1 - p_2| = d\} \quad . \quad (4)$$

The contrast of a gray image dominates human perception. If two images have different gray levels they still may be visually similar. Thus the correlogram method cannot be used directly.

Our modified correlogram algorithm works as follows. First, we sort the pixels in descending order. Then, we order the results of the preceding sorting by ascendant distances of pixels to the center of the image. The distance of a pixel to the image center is measured by the L2 distance. After sorting by gray value and distance to the image center, we select the top 20 percent of pixels and the gray values higher than a threshold to estimate the autocorrelogram histogram. We set the threshold zero in this task. Any two pixels have a distance, and we estimate the probability that the distance falls within an interval. The distance intervals we set are $\{(0,2), (2,4), (4,6), (6,8), (8,12), (12,16), (16,26), (26,36), (36,46), (46,56), (56,76), (76,100)\}$. We calculate the probability of each interval to form the correlogram vector.

2.5 Gabor texture features

Gabor filter is widely adopted to extract texture features from images for image retrieval [Manjunath 96], and has been shown to be very efficient. Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energy of the signal. Texture features can then be extracted from this group of energy distributions. The scale (frequency) and orientation tunable property of Gabor filter makes it especially useful for texture analysis.

The Gabor wavelet transformation W_{mn} of Image $I(x,y)$ derived from Gabor filters according to [Manjunath 96] is defined in Eq. (5)

$$W_{mn}(x, y) = \int I(x, y) g_{mn}^*(x - x_1, y - y_1) dx_1 dy_1 \quad . \quad (5)$$

The mean μ_{mn} and standard deviation σ_{mn} of the magnitude $|W_{mn}|$ are used for the feature vector, as shown in Eq. (6).

$$\mu_{mn}(x, y) = \iint |W_{mn}(x, y)| dx dy, \text{ and } \delta_{mn} = \sqrt{\iint (|W_{mn}(x, y)| - \mu_{mn})^2 dx dy} \quad (6)$$

This image feature is constructed by μ_{mn} and σ_{mn} of different scales and orientations. Our experiment uses four (S=4) as the scale and six (K=6) as the orientation to construct a 48 feature vectors \vec{f} , as shown in Eq. (7).

$$\bar{f} = [u_{00}, \delta_{00}, u_{01}, \delta_{01}, \dots, u_{35}, \delta_{35}]. \quad (7)$$

3. Classification method

In this work, we use Support Vector Machine (SVM) [Boser 92] to learn image feature characteristics. SVM is an effective classification method. Its basic idea is to map data into a high dimensional space and find a separating hyperplane with the maximal margin. Given a training set of instance-label pairs (x_i, y_i) , $i=1, \dots, k$ where $x_i \in \mathcal{R}^n$ and $y_i \in \{1, -1\}^k$, the support vector machine optimizes the solution of the following problem:

$$\text{Min}_{w, b, \phi} \left(\frac{1}{2} w^T w + C \sum_{i=1}^k \phi_i \right) \text{ and } y_i (w^T \psi(x_i) + b) \geq 1 - \phi_i, \phi_i \geq 0 \quad (8)$$

Training vectors x_i are mapped into a higher dimensional space by the function ψ . Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. $C > 0$ is the penalty parameter of the error term. Furthermore, $K(x_i, x_j) \equiv \psi(x_i)^T \psi(x_j)$ is called the kernel function. In this paper we use LIBSVM [Lin 01] to classify the training data with a radial basis function or a polynomial function as the kernel function. The radial basis function (RBF) and the polynomial function used is defined in Eq. (8) and Eq. (9), respectively, where γ , r , and d are kernel parameters.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0. \quad (9)$$

$$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0. \quad (10)$$

4. Submissions to the ImageCLEF 2005 Evaluation

In the ImageCLEF 2005 Automatic Annotation Task, we have submitted five SVM-based runs. Table 1 gives an overview of the features and the error rates of submit runs. The error rate is noted that each .1% corresponds to 1 misclassification, because this task has a total of one thousand images to be classified. The first run, Façade scale feature with 64-feature vectors is used and the radial basis function is chosen as the kernel function of SVM. Both of the second run and the third run use all 324 features, but they use different kernel functions for the SVM. The fourth run uses two kind of features, Façade scale and fuzzy histogram layout, and contains 208 features. The fifth run uses the coherence moment feature only and the radial basis kernel function for SVM.

In the image features used in our experiment, the facade scale feature that directly scales down an image contains the most spatial relationship. The fuzzy histogram layout feature divides an image into nine sections, which result in less spatial information than the facade scale feature; however, this feature is more invariant in image shift. The coherence moment considers the image rotation and shift, but cannot carry much spatial information.

In our experiments, the first run has the best result, with error rate 24.7%. The second run, which has error rate 24.9%, uses all image features but does not have better result than the first run, because the first run contain the most spatial information. Others image features do not improve the description of spatial relationship. In medical image data, the spatial distribution of pixels is the most significant. The fifth run contains the least spatial information, thus it has the worst result.

In the ImageCLEF 2005, one experiment for a nearest neighbor classifier that scales down images to 32*32 pixels and use the Euclidean distance for comparison has error rate 36.8%, which means that 368 images are misclassified. This experiment uses a feature very similar to the façade features; however the facade image feature scales down an image to only 8 x 8 pixels., It can be observed that the representation of the facade image feature is more concise and has better result than the 32x32-pixel features.; furthermore, the SVM method has better performance than the Euclidean distance metric.

Table 1: Features for the different submissions and the evaluation result.

Submission runs	Image features	SVM kernel function	error rate %
nctu_mc_result_1.txt	Facade scale: 64 vectors	radial basis function	24.7
nctu_mc_result_2.txt	All features: 324 vectors	radial basis function	24.9
nctu_mc_result_3.txt	All features: 324 vectors	polynomial	31.8
nctu_mc_result_4.txt	Facade scale, Histogram layout: 208 vectors	radial basis function	28.5
nctu_mc_result_5.txt	Coherence Moment: 56 vectors	radial basis function	33.8

5. Conclusions and future work

In this paper, several image features are examined for medical image data. The medical image application is unlike general-propose images. In general propose images, the representation always consider the invariance in image rotation, zooming and shift. Medical images have more stable camera settings than general propose images; therefore, the spatial information becomes very important in medical images, and we must improve the representation regarding spatial relation in this kind of images.

We use the support vector machine as a classifier; it is very efficient, but it seems that the SVM lacks the ability of feature selection. The fourth run also contains Facade scale feature but the result is worse than the first run. In the future, we plan to develop the feature selection technology for the SVM to improve the performance.

Reference

- [Manjunath 96] B. S. Manjunath and W. Y. Ma. "Texture features for browsing and retrieval of large image data" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (Special Issue on Digital Libraries), Vol. 18 (8), August 1996, pp. 837-842.
- [Flickner95] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, Query by Image and Video Content: The QBIC system, *IEEE Computer* 28 (9) (1995) 23-32.
- [Carson 99] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, Blobworld: A system for region-based image indexing and retrieval, in: D. P. Huijsmans, A. W. M. Smeulders (Eds.), *Third International Conference On Visual Information Systems (VISUAL' 99)*, no. 1614 in Lecture Notes in Computer Science, Springer-Verlag, Amsterdam, The Netherlands, 1999, pp. 509-516.
- [Belongie 98] S. Belongie, C. Carson, H. Greenspan, and J. Malik, Color and texture based image segmentation using EM and its application to content-based image retrieval, in: *Proceedings of the International Conference on Computer Vision (ICCV'98)*, Bombay, India, 1998, pp. 675-682.
- [Squire 99] D. M. Squire, W. Muller, H. Muller, and J. 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*, Kangerlussuaq, Greenland, pages 143-149, June 1999.
- [Wang 01] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 23(9):947-963, September 2001.
- [Han01] J., Han, M., Kamber,; Data Mining: Concepts and Techniques. *Academic Press*, San Diego, CA, USA (2001)
- [Keysers04] D. Keysers, W. Macherey, H. Ney, and J. Dahmen. Adaptation in Statistical Pattern Recognition using Tangent Vectors. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26(2):269-274, February 2004.
- [Swain91] Swain M.J. and D. H. Ballard, "Color Indexing", *International Journal of Computer Vision*, Vol. 7, pp.11-32, 1991.
- [Cheng 04] Pei-Cheng Cheng, Been-Chian Chien, Hao-Ren Ke, and Wei-Pang Yang, "KIDS's evaluation in medical image retrieval task at ImageCLEF 2004", Working Notes for the CLEF 2004 Workshop September, Bath,UK , pp. 585-593.
- [Boser 92] Boser, B., I. Guyon, and V. Vapnik (1992). A training algorithm for optimal margin classifiers. *In Proceedings of the Fifth Annual Workshop on Computational Learning Theory*.
- [Lin 01] Chang, C.-C. and C.-J. Lin (2001). LIBSVM: a library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.