Classification-Driven Medical Image Retrieval

Yanxi Liu and **Frank Dellaert**
{yanxi,dellaert}@cs.cmu.edu
The Robotics Institute, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213

ABSTRACT

We propose a novel image retrieval framework centered around classification-driven search for a weighted similarity metric for image retrieval. This approach is firmly rooted in Bayes decision theory. Given a well-defined image set, we argue that image classification and image retrieval share fundamentally the same goal. Thus, the distance metric defining a classifier that performs well on the image set should also generate good results when used as the similarity metric for image retrieval. In this paper we report our methodology and initial results on neuroradiological image retrieval, where the approximate bilateral symmetry of normal human brains is exploited.

1 Introduction

Existing "content-based" image retrieval systems depend on general visual properties such as color and texture to classify diverse, two-dimensional (2D) images [4, 12, 15, 11, 10, 13]. These general visual cues often fail to be effective discriminators for image sets taken within a single domain, where images have subtle, domain-specific differences. Furthermore, global statistical color and texture measures do not necessarily reflect the meaning of an image.

Databases composed of 3D volumetric or 2D images and their collateral information in a particular medical domain provide simple, well-defined training sets, where the meaning of each image is the pathological classification indicated by that image. By **well-defined image set** S, we mean a finite number of image classes C are defined on an image set S, any pair of elements in C is mutually exclusive, and each image has one class only. During system performance evaluation, the retrieval results are readily quantified by comparing the pathology classes in the query image with that of the retrieved images. As a test-bed, we have chosen a multimedia digital database in human neurology from the National Medical Practice Knowledge Bank project [6]. This is a database containing a large set of medical cases in different sub-branches of medicine, with a current focus on neurology. Each case in the database is composed of at least one 3D image, either Computed Tomography (CT) or Magnetic Resonance images (MRI).

2 Our Approach

2.1 Basic Idea

Given a well-defined image set, image classification and image retrieval share fundamentally the same goal. Our hypothesis is that *a distance metric* used by a classifier that gives good classification results should also be expected to provide good image retrieval results when used as the image similarity metric for retrieval. Therefore, in our work image classification is used as an image index feature selection tool to find a good similarity metric for image retrieval. The net effect is equivalent to warping the original multidimensional feature space such that the image points corresponding to the same class are well-clustered in the warped feature space. It should then be obvious that image retrieval using nearest neighbors in this feature space can achieve good results.

2.2 Basic Framework

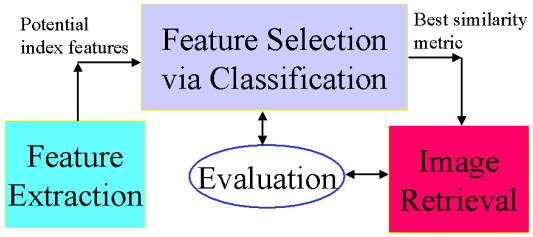

Figure 1: Three steps in classification-driven image retrieval

We propose an image retrieval framework consisting of three stages as shown in Figure 1: 1) **feature extraction**: a pool of candidate image indexing features is extracted, 2) **feature selection** via image classification: determine, quantitatively via classification, the most discriminating subset of weighted features across predefined image classes, and 3) **image retrieval** using k nearest neighbors in the selected feature space: the weighted subset of the features that has the best performance in classification is used as a similarity metric for image retrieval. This distinguishes our work from most existing systems, where no direct evaluation of the chosen image features is performed. A typical content-based image retrieval system simply extracts certain predetermined image features from both the

query image and the images in the database, finds the nearest N neighbors of the query image in the feature space and returns those images as the retrieved result. The difficult feature weighting problem is usually left for the human user to decide. Besides providing no insights to the distributions of the database image feature space in terms of image classes, this user-defined-feature-weighting method is not always a manageable task for the user, especially when the user is not experienced or when multiple indexing features ($>> 5$) are used.

What is missing in the common image retrieval practice is an evaluation of the extracted image features **before** they are used for image retrieval. Critical questions are left unanswered, such as "*Are these features sufficient to capture the content of the image? Is there any redundancy in these features for discriminating different images? Which subset of these features has the most discriminating power across image semantic classes? and what are their relative weights?*". In contrast, the method we propose directly answers these questions. In summary, our approach uses memory-based learning to explore, discover and manipulate image feature space, in the hope of finding the most direct, effective and economical mapping from a proper subset of non-uniformly rescaled image features to their corresponding image classes.

2.3 Bayesian Classifiers

A memory based learning (MBL) technique, called *Kernel regression* (KR), is applied to classify images in the database [1, 9]. Recall that the posterior probability $P(c|\mathbf{x})$ of an image being in class c when feature $\mathbf{x}$ is observed can be computed via Bayes law [3]:

$$P(c|\mathbf{x}) = \frac{P(\mathbf{x}|c)P(c)}{P(\mathbf{x}|c)P(c) + P(\mathbf{x}|\bar{c})P(\bar{c})} \qquad (1)$$

The *prior probability* $P(c)$ of a class c can be estimated easily from labeled training data by dividing N_c, the number of instances in class c, by the total number of instances N in the training set: $P(c) \approx N_c/N$. The conditional density $P(\mathbf{x}|c)$ for each class is estimated using the Parzen window [3].

Changing the kernel width σ in KR density estimation or the distance metric (feature weighting) yields a whole space of classifiers. Our approach consists of an off-line, combinatorial search in the space of classifiers to find one that minimizes a classification performance measure: cross-entropy. *Cross-entropy* is defined as the negative log-likelihood of the training data given a specific classifier, and is therefore a measure for how well a classifier performs. Minimizing this cross-entropy will yield a metric and a kernel width σ for which kernel regression best approximates the a posterior probabilities $P(c|\mathbf{x})$, and is thus optimally suited for classification [2]. To prevent overfitting to the training data, each classifier is evaluated using leave-one-out cross-validation[8] on a training set, containing roughly two-thirds of the available data in the database. The rest of the data is set aside for evaluation purposes, and it is on this set our results are reported.

The main result of the search is a similarity metric, i.e. a weighted set of the most discriminating features for use in a classifier. The classification process screens out the majority of the features that were extracted for each image, and ends up with a proper subset of the original feature attributes. If large redundancy exists in the initial feature set then one should expect several feature subsets that have quantitatively equivalent discriminating power to be found. In that case, which individual feature is included in each subset, or even which subset is used by the classifier, is not important.

2.4 Classification Evaluation

In general, a classification will never be perfect, since the densities of classes overlap in feature space. This motivates the use of decision theory [3] to minimize the cost of deciding on a particular class for a given sample. Specifically, in the domain of medical image classification/retrieval, it is imperative to minimize the *false negative rate* **FNR**, i.e. minimize the occurrence of pathological cases being classified as normal. This motivates a *cost matrix* structure (Figure 2). A *false negative penalty* $w > 1$ is incurred whenever a pathological image is classified as normal, whereas a unit cost or zero cost is incurred when a normal image is classified as pathological or when a class chosen is the correct class, respectively. Thus, given three classes: 1. normal, 2. blood and 3. stroke, we typically use a cost matrix of the form shown in Figure 2. A classification decision is made by minimizing the expected risk associated with choosing a class. For example, $R1$ in Figure 2 is the risk to take when choosing class 1, i.e. normal, for the given image. From the expression one can see that the higher the value of w, the higher the risk. Thus, by increasing w, false negative decisions are effectively punished.

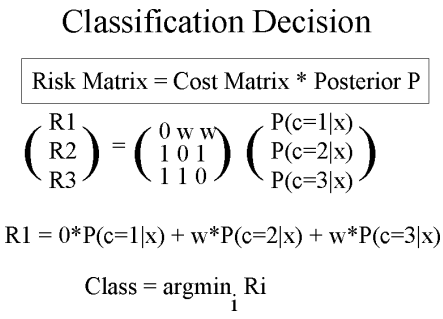

Figure 2: Risk matrix. R1 is the risk to take when choosing class 1, i.e. normal. The higher the value of w, the higher the risk. By increasing w, false negative decisions are effectively punished.

2.5 Image Retrieval and Evaluation

Once a locally optimal classifier and associated distance metric have been found, we apply this same metric to the image retrieval problem. Assume the extracted image features form an N dimensional vector space, the overall distance function D of two images a, b is defined as

$$D(a,b) = \sqrt{(\vec{A}-\vec{B})^T \Sigma^{-1} (\vec{A}-\vec{B})} \qquad (2)$$

where $\vec{A} = f(a), \vec{B} = f(b)$ are the N dimensional image feature vectors of images a and b respectively, Σ is an $N \times N$ covariance matrix. In this paper we use non-uniformly scaled Euclidean distance. Thus Σ is a diagonal matrix, whose diagonal elements are the feature weights being discovered during the classifier selection process described above.

The image retrieval is done using K nearest neighbor (KNN) as commonly done in image content-based retrieval. The evaluation of retrieved images can be simply the true positive rate or *precision* in the top K retrieved images. Let C_p be the number of correct pathology class in the top K retrieved images as that of the query image, then the top K precision rate is defined as $\mathbf{TPR_K} = C_p/K$.

3 Experiments

A study was performed using a data set of 48 volumetric CT brain scans containing normal (26), stroke (14) and blood cases (8). These are clinical CT images collected directly from a local hospital emergency room. These images bear a close resemblance to the images in appearance that will be used as query images.

A 3D medical image is usually taken as a stack of parallel 2D images. Instead of working directly with the 3D images as the descriptive units in the database, 2D slices are used. 2D image are easier to display, and physicians primarily use 2D images as queries to access digital databases. It is also useful to identify normal slices in an otherwise pathological 3D brain image. Finally, it is hard to obtain a sufficient number of 3D images, whereas a small number of 3D images yields a large set of 2D slices. In fact, we go one step further and ultimately use *2D half slices* as the basic descriptive unit for testing our approach, doubling the amount of data at our disposal. This is justified under the assumption that normal human brains are approximately symmetrical, thus each half of a brain slice is potentially equivalent to the other half. The violation of this symmetry is an implication of pathology.

The total 3D image set S_{3D} is then divided into a training set containing two thirds of S_{3D} and a hold-out test set containing one third of S_{3D}. Care is taken to separate the test set in such a way that there is no mixing of 3D images between training and test set at the level of 2D slices. Instead, the test set is a completely separate set of 3D images, and is only used for evaluation purposes.

3.1 Image Preprocessing

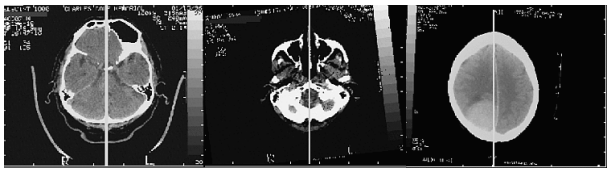

Figure 3: The ideal symmetry axis is extracted as the intersection of a 3D brain image and its midsagittal plane. The method works even in the presence of obvious asymmetries, for example the sinus near the frontal lobe in the left image.

The orientation of an imaging coordinate system often differs from the ideal head coordinate system $K_{X_0Y_0Z_0}$ by three rotation angles, *pitch, roll* and *yaw*, about X_0, Y_0 and Z_0 axes of $K_{X_0Y_0Z_0}$, respectively. Thus the first step in *image preprocessing* is to align all 3D images such that the *midsagittal* planes of different brain images are parallel and pitch angle is zero. A maximization of mutual information affine registration algorithm [7] and a 2D-to-3D cross correlation registration algorithm [5] are used to minimize rotational errors among 3D images. The effect of midsagittal extraction is not to find where the midsagittal plane is, but where it is supposed to be. This is especially true for pathology brains since the midsagittal plane is often distorted (shifted or bended) due to large lesions. Figure 3 shows some sample results after the midsagittal plane is extracted and intersected with 2D slices. After image preprocessing, each of the half slices is labeled according to its pathology class with the help of a neuroradiologist.

3.2 Feature Extraction

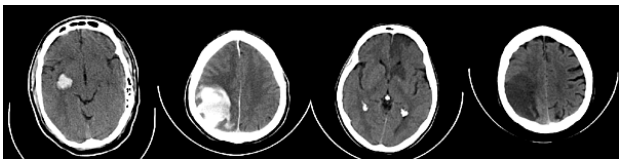

Figure 4: Variations of lesions.From left to right: acute blood, acute blood, infarct, infarct

In most medical image analysis, segmentation of lesion(s) in an image is a necessary first step. However, it is well known that lesion segmentation can be a very hard problem, particularly since each 3D image can contain multiple lesions, and each lesion takes no *a priori* form, density value, shape or geometric and anatomic location (see Figure 4). Many image retrieval systems depend on human experts to hand pick the lesion as the initial input.

We take a completely different approach: instead of trying to precisely locate the lesion, we collect a set of relatively simple and computationally inexpensive image features based on the assumption of normal brain

symmetry. The hypothesis is that some (appropriately weighted) subset of these simple features, collectively, may capture the statistical 3D appearance model of the pathology accurately. If none of these selected features, or any subsets of them, are effective, the result of the classification will reflect that fact. Only then should other, perhaps more expensive means, be used to extract more relevant features.

To extract features that will quantitatively describe the amount and type of asymmetry in the image, we use three types of symmetry-descriptive features: 1) global statistical properties, 2) measures of asymmetry of half brains, and 3) local asymmetrical region-based properties. In addition, some global features are also used, such as the mean of the whole image and the number of edge pixels in each quarter (top-left, top-right, bottom-left, bottom-right) of a Canny-edged image. These features are extracted from the original image I with its midsagittal plane centered vertically in the middle of I, the difference image D of the original image and its mirror reflection with respect to $X_0 = 0$, the thresholded difference image, and the original image **masked** by the thresholded binary image (Figure 5).

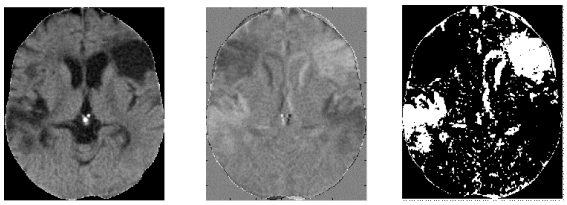

Figure 5: Left: original image I with midsagittal plane aligned. Middle: difference image $D = I - reflect_v(I)$. Right: thresholded D

Global statistical properties are computed over the entire half brain, and include features like the mean and standard deviation of grey-level intensity. *Asymmetry features* are obtained by voxel-wise comparison of corresponding left and right half brains. Two techniques are used to obtain these features. (1) The first is to simply subtract out a vertically reflected image of the left brain from its corresponding right half, to obtain a difference image D (Figure 5). Asymmetries show up as large positive or negative density values in the difference image. Numeric features are then computed as counts of how many voxels remain after thresholding D with different values. (2) In the second technique, each image is smoothed with a Gaussian filter having a standard deviation of 5, 9, or 15 pixels respectively. This is equivalent to examining images at different resolutions. Then the difference between each voxel and its counterpart in the opposite symmetrical half is recorded. If that counterpart voxel falls significantly outside the estimated image gray-value distribution, i.e. the difference is greater than 3, 4, or 5 standard deviations, it is flagged as being significantly different, and again the number of voxels that pass this threshold test are counted. Finally, a set of *local region statistics* are generated by masking the original image with the threshold images obtained in the previous step. Intensity means and variances are then computed over these asymmetrical areas, and thus pertain only to the local areas where asymmetries are present. The feature extraction is carried out on 2D half slices, yielding a total of 1250 labeled data points in a 48-dimensional feature space.

3.3 Searching the Space of Classifiers

To look for a distance metric in feature space having locally optimal classification performance we used a proprietary combinatorial search engine called "Vizier", implemented at CMU by Moore et al. [14]. A classifier is defined by a distance metric and a smoothing parameter σ. We look for a kernel regression classifier that minimizes the leave-one-out cross-entropy of the training data given the classifier. The "Vizier" search engine works by searching through a large set of possible classifiers within the user's specifications, and the search stops when either it exhausts all the possible choices or a time limit given by the user is reached. Vizier applies a number of combinatorial search algorithms similar to multiple restart local search, ranging from hill-climbing in metric space to standard feature selection searches common in the machine learning literature.

Typically, we run the search for a full hour on standard high-performance workstations. The output from Vizier is a specification of the best similarity metric found so far, consisting of the weight on each input attribute (a warped/scaled Euclidean space which is a proper subspace of the original feature space in terms of dimensions) and the smoothing parameter σ.

3.4 Evaluation of Classification Results

Since the search engine yields an "anytime" best classifier and contains some non-determinism, the exact classifier found may vary from run to run. However, a quantitative measure is used describing its improvement over the prediction of the most common class made by the found classifier. This score indicates the discriminating power of the selected features. Different combinations of features can yield the same discriminative power because of the redundancy between features. One typical run found a classifier using 9 features out of the original 48, all equally weighted in the distance metric. A list of the specific features is shown below (given a 2D slice I, its left or right half IH, the difference image $D = I - fliplr(I)$, and the right or left half of the difference image DH):

1. the sequence labeling of 2D slices
2. the mean absolute value of the X direction gradient of IH
3. the ratio of non-zero pixels remaining in D when thresholded at 20/255

4. white60 – the ratio of non-zero pixels remaining in D when thresholded at 60/255

5. the ratio of non-zero pixels remaining in DH when thresholded at 60/255

6. after Gausian smoothing of image I with $\sigma = 5$ and computing a new difference image D of I, the mean of non-zero pixels remaining in DH when thresholded at 5σ

7. after Gausian smoothing of image I with $\sigma = 15$ and computing a new difference image D, the mean of non-zero pixels remaining in D when thresholded at 3σ

8. after canny-edge the image I with threshold 15 and $\sigma = 4$, the total number of white pixels in the top-left quarter

9. let A be the sum of a pixel-wise multiplication of the absolute value of DH and the absolute value of vertical gradient of IH, and B be the sum of the absolute values of DH, then A/B.

Figure 6 displays the ROC curves (true positive rate / false positive rate) of the chosen classifier on classifying 2D half-slices, both for the training data and the separate test set. As can be seen from the figure, the performance on the training set is much better than the performance on the unseen test-set, which is to be expected, as we searched to maximize the former. Thus, reporting performance figures on the training data will give optimistic information on the performance of the resulting classifiers. The use of the cost-matrix and such summary graphs provide a good quantitative overview of the strengths and weaknesses of a particular set of selected features. It can also be used to select an appropriate w value to use when classifying new, unlabeled images.

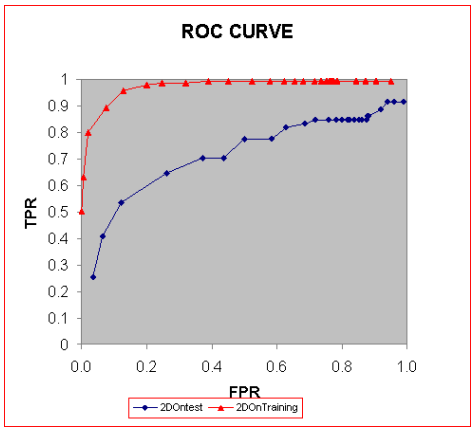

Figure 6: ROC curve for 2D classification on training images using LOO, and on a hold-out testing image set.

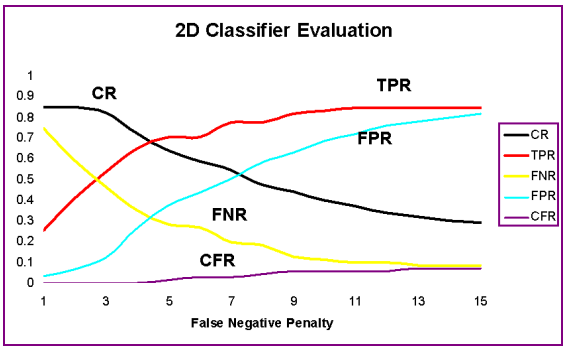

Figure 7: 2D classification rates for increasing value of w. CR: classification rate, TPR: true positive rate, FPR: false positive rate, FNR: false negative rate, CFR: confusion (on pathologies) rate.

3.5 Possible Extensions

3.5.1 Layered Classifiers

It is possible to construct a layered classifier on top of the 2D image classifiers. For example, one can construct a classifier to determine the pathology class of a 3D image by using the output of the 2D classifier as the input data. For a given 3D image with n 2D half slices, n_N, n_S, n_B are the numbers of predicted normal, stroke and blood classes, where $n = n_N + n_S + n_B$. Using ratios of $r_S = n_S/n$ and $r_B = n_B/n$, the image distribution in this two-dimensional feature space can be observed in Figure 8. From these feature spaces,

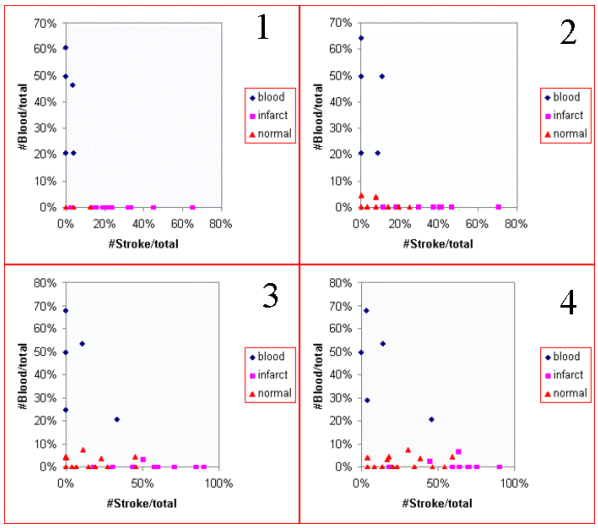

Figure 8: The image points migrate in the two-ratio space while increasing the false negative penalty value of the 2D three-class classifier.

one can observe that separating blood cases from normal is easy but it is not true for the stroke cases, especially when the false penalty w for 2D classifiers is

increased. Though Figure 8 only demonstrates a simple case, it is conceivable that for more complicated distributions one can apply the same approach as we have done with the 2D classifier on a new feature space to find an optimal classifier. When we tried this using the two ratios alone, a classifier with 50% strength on feature r_S and full strength on feature r_B is found, its true positive rate is evaluated at above 80% on the 3D image test set.

3.5.2 Query-Dependent Similarity Metric

Our approach is ideally suited to tuning a similarity metric to the specific type of query that may be submitted to the system. As an example, one might only be interested in finding whether an image shows a pathology or not, without regard for the type of pathology. In this case, the metric that classifies the images into normal/pathological might be different from one that needs to make the additional distinction between stroke and blood. When we tried this, the locally optimal classifier found a different metric than the one from the ternary classification case, however, they give quantitatively comparable classification rates. This result indicates that the features we used have equivalent discriminating power over the 3-class and the 2-class problems, but one can expect better performance when the query is "does this image belong to the blood class?" due to the clear class separation between blood class and the rest of the classes in the feature space.

3.6 Image Retrieval

The similarity metric found in image classification can now serve as an image index vector for retrieving images by finding the nearest neighbors in the feature space, as is conventionally done in content-based image retrieval. However, the dimension of the index feature vector is now much reduced with respect to the original feature space.

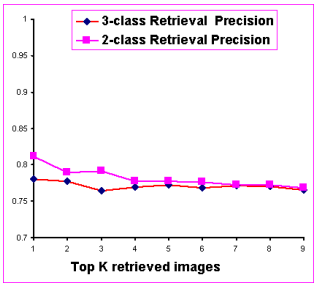

Figure 9: The mean value of retrieval precision as a function of the top K retrieved images.

Figure 9 shows the mean retrieval precision rate for all the hold out test images, one for the three-class case and one for the two-class case described in the previous section. The two different optimal metrics found during classification are used here as the similarity metric for retrieval. One can observe a slightly better performance for the 2-class than for the 3-class image set. This is to be expected since 2-class classification leaves less space for errors to be made than the 3-class classification problem. However, in this problem the difficult class separation is between normals and strokes, which can be observed from the original feature values and 2D classifier output (Figure 8).

Given near 80% precision rate on average, this result implies that in average 8 out of 10 top ranked retrieved images have the same pathology as the query image. Figures 10 and 11 show two of the best retrieved results for pathology cases.

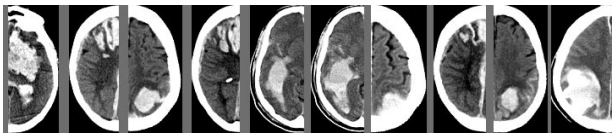

Figure 10: Left most: query image with acute blood. The first nine retrieved half slices follow, from left to right in descending order of similarity. The pathologies in the retrieved images are all acute blood.

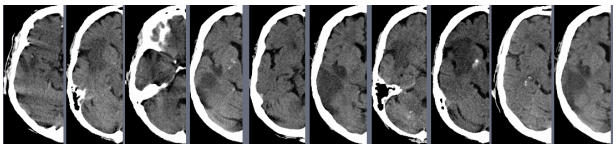

Figure 11: Left most: query image with stroke (infarct). The first nine retrieved half slices follow,from left to right in descending order of similarity. The pathologies in the retrieved images are: infarct, infarct, infarct, normal, infarct, infarct, infarct, normal, infarct.

4 Conclusion and Future Work

From our preliminary study, two stages with five essential components have been identified for constructing a classification-driven image retrieval system (Figure 12). The two stages are (A) off-line similarity metric learning (feature selection) and (B) on-line image retrieval. The five essential components are (1) Image preprocessing, (2) Image feature extraction, (3) Feature selection via image classification, (4) Image retrieval, and (5) Quantitative evaluation. Though the two stages share some common components the goals and constraints differ. In stage (A) the goal is to find the best and smallest subset of image features that capture image classes. It requires an explicitly labeled image database, and extensive search on a large feature attribute matrix. High computational speed is a plus but not necessary. Stage (B), on the other hand, demands fast retrieval speed and presentation of retrieved images in addition to a given similarity metric and a query image.

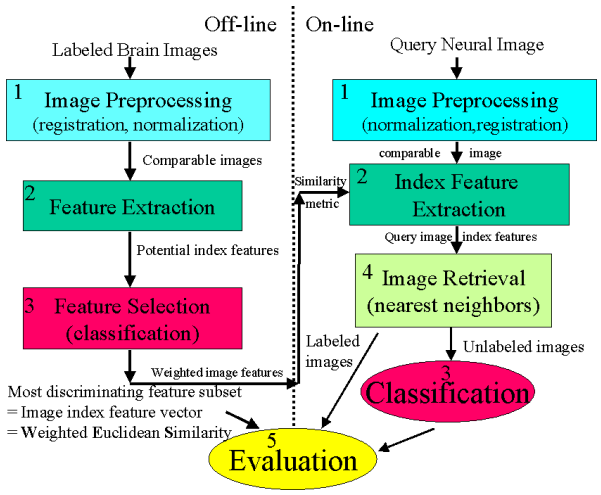

Figure 12: A basic framework and individual components for classification-driven image retrieval

In any application domain, coming up with the potential feature set is still an important and not easily automated step. Candidate features need to be selected using considerable domain knowledge. One advantage of classification-driven approach is that using the classification as a feature screening tool with the quantitative measurements on classification result, the effort on formating domain specific features can be under control (when more domain knowledge is called for), or can even be reduced (when to stop). Other advantages include the capability to access unlabeled image databases, i.e. an extra step can be added to classify retrieved un-labeled images (Figure 12).

Future work includes the study of different sets of most discriminative features for different purposes or on different subsets of the image database, and dynamically switching from one similarity metric to another during retrieval. We expect the similarity metrics to vary accordingly, as we have observed in experiments with binary and ternary classes reported in this paper. We also would like to combine visual features with collateral information such as age, sex, and symptoms of the patient to obtain a better retrieval rate and faster retrieval speed. Other kind of features such as eigenfeatures can also be added. The basic framework presented here has provided us with such an *information fusion* capability.

5 Acknowledgement

The authors would like to thank Dr. Rothfus of Radiologic Sciences, Allegheny General Hospital for his medical guidance, Prof. Andrew Moore and Dr. Jeff Schneider of CMU for helpful discussions. This research is partly supported by the Allegheny-Singer Research Institute under prime contract through the National Institute of Standards and Technology (NIST#70NANB5H1183).

References

[1] C. Atkeson, S. Schaal, and Andrew Moore. Locally weighted learning. *AI Review*, 11:11–73, 1997.

[2] C. M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, 1995. ISBN:0198538499.

[3] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, 1973.

[4] D. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. *Journal of Intelligent Information Systems*, 1994.

[5] Y. Liu, R.T. Collins, and W.E. Rothfus. Automatic Bilateral Symmetry (Midsagittal) Plane Extraction from Pathological 3D Neuroradiological Images. *SPIE's International Symposium on Medical Imaging 1998*, 3338(161), February 1998.

[6] Y. Liu, W.E. Rothfus, and T. Kanade. Content-based 3d neuroradiologic image retrieval: Preliminary results. *IEEE workshop on Content-based accessof Image and Video Databases in conjunction with ICCV'98*, January 1998.

[7] F. Maes, A. Collignon, D. Vandermeulun, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187,198, 1997.

[8] A. W. Moore, D. J. Hill, and M. P. Johnson. An Empirical Investigation of Brute Force to choose Features, Smoothers and Function Approximators. In S. Hanson, S. Judd, and T. Petsche, editors, *Computational Learning Theory and Natural Learning Systems, Volume 3*. MIT Press, 1994.

[9] A. W. Moore, J. Schneider, and K. Deng. Efficient locally weighted polynomial regression predictions. In *Proceedings of the 1997 International Machine Learning Conference*. Morgan Kaufmann, 1997.

[10] Virage Inc. Home page. http://www.virage.com.

[11] A. Pentland, R.W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *IJCV*, 18(3):233–254, June 1996.

[12] R.W. Picard. A societry of models for video and image libraries. *IBM Systems Journal*, 1997.

[13] Y. Rui, T.S. Huang, and S. Mehrotra. Relevance feedback techniques in interactive content-based image retrieval. In *SPIE/IS&T Conf. on Storage and Retrieval for Image and Video Databases VI*. volume 3312, January 1998.

[14] J. Schneider and A.W. Moore. A locally weighted learning tutorial using vizier 1.0. 1997.

[15] M.J. Swain. Interactive indexing into image databases. *SPIE*, 1908, 1993.