# Content Based Image Retrieval for Medical Images Techniques and Storage Methods-Review Paper

Prof. K.Wanjale
Asst.Professor
VIIT,Kondhwa,Pune-411048
.

Tejas Borawake
Student
VIIT,Kondhwa,Pune-411048

Shashideep Chaudhari
Student
VIIT,Kondhwa,Pune-411048

## ABSTRACT
In the medical field, images, and especially digital images, are produced in ever increasing quantities and used for diagnostics and therapy. Content based access to medical images for supporting clinical decision making has been proposed that would ease the management of clinical data and scenarios for the integration of content-based access methods into Picture Archiving and Communication Systems (PACS) have been created. This article gives an overview of available literature in the field of content based access to medical image data and on the technologies used in the field. Section 1 gives an introduction into generic content based image retrieval and the technologies used. Section 2 describes the basic algorithm used in the implemented systems. Section 3 describes various methods of implementing CBIR. New research directions are being defied that can prove to This article also identifies explanations to some of the outlined problems in the field as it looks like many propositions for systems are made from the medical domain and research prototypes are developed in computer science departments using medical datasets. Still, there are very few systems that seem to be used in clinical practice.
Content-Based Image Retrieval (CBIR) has been a topic of research interest for nearly a decade. Approaches to date use image features for describing content. A survey of the literature shows that progress has been limited to prototype systems that make gross assumptions and approximations.

## General Terms
I.4.7 [Image processing and Computer Vision]: Feature measurement

## Keywords
CBIR: Content based image retrieval
GFD: Generic Fourier Descriptor
KD trees: K dimensional trees.

## 1. INTRODUCTION
In the medical field, images, and especially digital images, are produced in ever increasing quantities and used for diagnostics
and therapy. Content based access to medical images for supporting clinical decision making has been proposed that would ease the management of clinical data and scenarios for the integration of content-based access methods into Picture Archiving and Communication Systems (PACS) have been created. This article gives an overview of available literature in the field of content based access to medical image data and on the technologies used in the field. Section 1 gives an introduction into generic content based image retrieval and the technologies used. Section 2 describes the basic algorithm used in the implemented systems. Section 3 describes various methods of implementing CBIR. New research directions are being defied that can prove to be useful.

This article also identifies explanations to some of the outlined problems in the field as it looks like many propositions for systems are made from the medical domain and research prototypes are developed in computer science departments using
medical datasets. Still, there are very few systems that seem to be used in clinical practice.

Content-Based Image Retrieval (CBIR) has been a topic of research interest for nearly a decade. Approaches to date use image features for describing content. A survey of the literature
shows that progress has been limited to prototype systems that make gross assumptions and approximations.

## 2. Content Based Image Retrieval Systems
Content-based image retrieval hinges on the ability of the algorithms to extract pertinent image features and organize them
in a way that represents the image content. Additionally, the algorithms should be able to quantify the similarity between the query visual and the database candidate for the image content as perceived by the viewer. Thus, there is a systemic component to CBIR and a more challenging semantic component.
The proposed CBIR framework is shown in Figure 1. The images are kept in a database called Image Database. After preprocessing, images are segmented by using the method described. CBIR system consists of two major steps. The first one is the feature extraction, where a set of feature is generated to represent the content of each image. The second one is the similarity measurement where a distance between the query image and each image in the database is computed using their feature vectors. For all images in the database the features are extracted and stored in the database. Given a query image, such a system first extracts its feature vector and then compares it to those of the images stored in the database. The images of the database are ranked according to the distance of their feature vector to the query image.

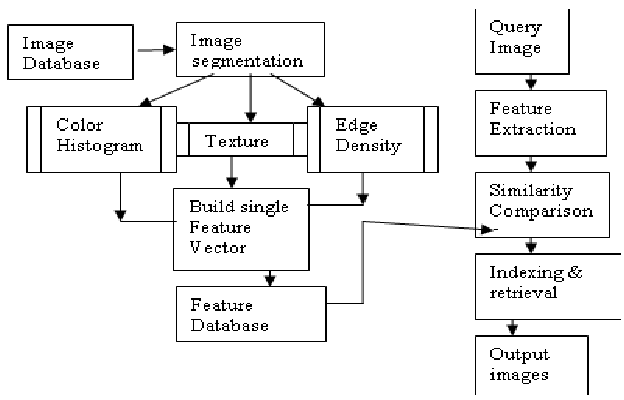


Fig. 1 CBIR framework

## 3. Methods used for implementing CBIR

### 3.1 Shape Based Method

For shape based image retrieval, the image feature extracted is usually an N dimensional feature vector which can be regarded as a point in a *N* dimensional space. Once images are indexed into the database using the extracted feature vectors, the retrieval of images is essentially the determination of similarity between the query image and the target images in database, which is essentially the determination of distance between the feature vectors representing the images. The desirable distance measure should reflect human perception. Various similarity measures
have been exploited in image retrieval. In our implementation we have used Euclidean distance for similarity measurement.

### 3.2 Texture Based Method

Texture measures have an even larger variety than color measures. Some common measures for capturing the texture of images are wavelets and Gabor filters where Gabor filters perform better and correspond well to. The texture measures try to capture the characteristics of images or image parts with respect to changes in certain directions and scale of changes. This is most useful for regions or images with homogeneous texture.

### 3.3 Continuous Feature Selection Method

This method deals with the "dimensionality curse" and the semantic gap problem. The method applies statistical association
rule mining to relate low-level features with high-level specialists knowledge about the image, in order to reduce the semantic gap existing between the image representation and interpretation. These rules are employed to weight the features according to their relevance. The dimensionality reduction is performed by discarding the irrelevant features (the ones whose weight are null). The obtained weights are used to calculate the similarity between the images during the content-based searching. Experiments performed show that the proposed method improves the query precision up to 38%. Moreover, the method is efficient, since the complexity of the query processing decreases along the dimensionality reduction of the feature vector.

### 3.4 With Automatically Extracted MeSH Terms

There is still a semantic gap between the low-level visual features(textures, colors) automatically extracted and the highlevel concepts that users normally search for (tumors, abnormal tissue)[2]. Proposed solutions to bridge this semantic gap are the connection of visual features to known textual labels of the images [3] or the training of a classifier based on known class labels and the use of the classifier on unknown cases [6]. Combinations of textual and visual features for medical image
retrieval have as of yet rarely been applied, although medical images in the electronic patient record or case databases basically always do have text attached to them. The complementary nature of text and visual image features for retrieval promises to lead to good retrieval results.

### 3.5 Using Low-Level Visual Features and

The image retrieval process consists of two main phases: preprocessing phase and retrieval phase. Both phases are described as follows.

The pre-processing phase is composed of two main components: a feature extraction model and a classification model. The input of the pre-processing phase is the original image database, i.e. images from the ImageCLEFmed collection, with more than 66,000 medical images. The output of the pre-processing phase is an index relating each image to its modality and a feature database.

The Feature Extraction Model

The feature extraction model operates on the image database to
produce two kinds of features: histogram features and metafeatures. Histogram features are used to build the feature database, which is used in the retrieval phase to rank similar images. Metafeatures are a set of histogram descriptors, which are used as the input to the classification model to be described later. Histogram features used in this system are :

- Gray scale and color histogram (Gray and RGB)
- Local Binary Partition histogram (LBP)
– Tamura texture histogram (Tamura)
– Sobel histogram (Sobel)

– Invariant feature histogram (Invariant) Meta features are calculated from histogram features in order to reduce the dimensionality.

These metafeatures are the four moments of the moment generating function (mean, deviation, skewness and kurtosis) and the entropy of the histogram. Each histogram has five associated metafeatures, meaning a total of 30 meta-features with information of color, texture, edges and invariants.

## 4. Storage and access methods

Image retrieval system is a huge system and it needs to store many images in database. That's why it requires large database to store the images. Common storages are being used are relational databases, inverted files, self made structures or simply keep entire index in the main memory which will inevitably cause problems when using large databases. These methods often need to use dimension reduction techniques or pruning methods to allow for an efficient and quick access to the data. Some indexing techniques such as KD trees. Principle of Component analysis (PCA)[7] for feature space reduction is being used. Another space reduction technique is Independent Component analysis (ICA).
Due to the lack of a common database for evaluation in CBIR with known relevances we use two databases where relevances are implicitly given by classifications. These databases are chosen as representatives for two different types of CBIR tasks: The WANG database represents an CBIR task with arbitrary photographs. In contrast, the IRMA database represents a CBIR task in which the images involved depict more clearly defined objects, i.e. the domain is considerably narrower.

### 4.1 WANG

The WANG database[4] is a subset of 1000 images of the Corel database which were selected manually to form 10 classes of 100 images each. The images are subdivided into 10 sufficiently distinct classes (e.g. „Africa., „beach., „monuments., „food.) such that it can be assumed that a user wants to find the other images from the class if the query is from one of these ten classes. This database was created at the Pennsylvania State
University and is publicly available.

## 4.2 IRMA

The IRMA database[4] is a database of 1617 medical radiographs collected in a collaboration project of the RWTH Aachen University. The complete data is labelled using a multiaxial code describing several properties of the images. The data were divided into the six classes abdomen, skull,limbs., „chest., „breast. and „spine., describing different body regions. The images are of varying sizes

## 5. REFERENCES

[1] Content-based Retrieval of Medical Images by ContinuousFeature Selection Pedro H. Bugatti, Marcela X. Ribeiro, Agma J. M. Traina, Caetano Traina Jr Department of Computer Science – ICMC, University of Sao Paulo at Sao Carlos – USP.

[2] Enriching content-based medical image retrieval with automatically extracted MeSH-terms by Müller H, Ruch P,Geissbuhler A.

[3] Content-Based Medical Image Retrieval Using Low-Level Visual Features and Modality Identification by Juan C. Caicedo, Fabio A. Gonzalez and Eduardo Romero.

[4] Features for Image Retrieval: A Quantitative Comparison by Thomas Deselaers, Daniel Keysers, and Hermann Ney.

[5] Content Based Image Retrieval For Medical Images Using Generic Fourier Descriptor by A. Grace Selvarani1 and Dr. S. Annadurai2.

[6] Image Retrieval: Ideas, Influences, and Trends of the New Age Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. The Pennsylvania State University.

[7] A Review of Content–Based Image Retrieval Systems in Medical Applications – Clinical Benefits and Future Directions Henning Miller, Nicolas Michoux, David Bandon and Antoine Geissbuhler Division for Medical Informatics, University Hospital of Geneva Rue Micheli–du–Crest 21, 1211 Geneva 14, Switzerland.

[8] A Universal Model for Content-Based Image Retrieval S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak.