Privacy Preserved CBIR Algorithm based on Feature Integration

Nitish Bhattacharjee

Department of Computer Science & Engineering
Indian Institute of Technology (Indian School of Mines), Dhanbad

May 1, 2019

Contents

The Basic Idea

- Content based image retrieval(CBIR), performs retrieval of the images from a large db, that are similar to an image given as query.
- CBIR has its application in domains such as medical report images, weather forecasting, historical research etc.
- Content refers to the information present in images which can be either texture, shape or color.
- The risk in CBIR is of privacy leakage of sensitive images. Hence images need to be encrypted before being stored or outsourced.

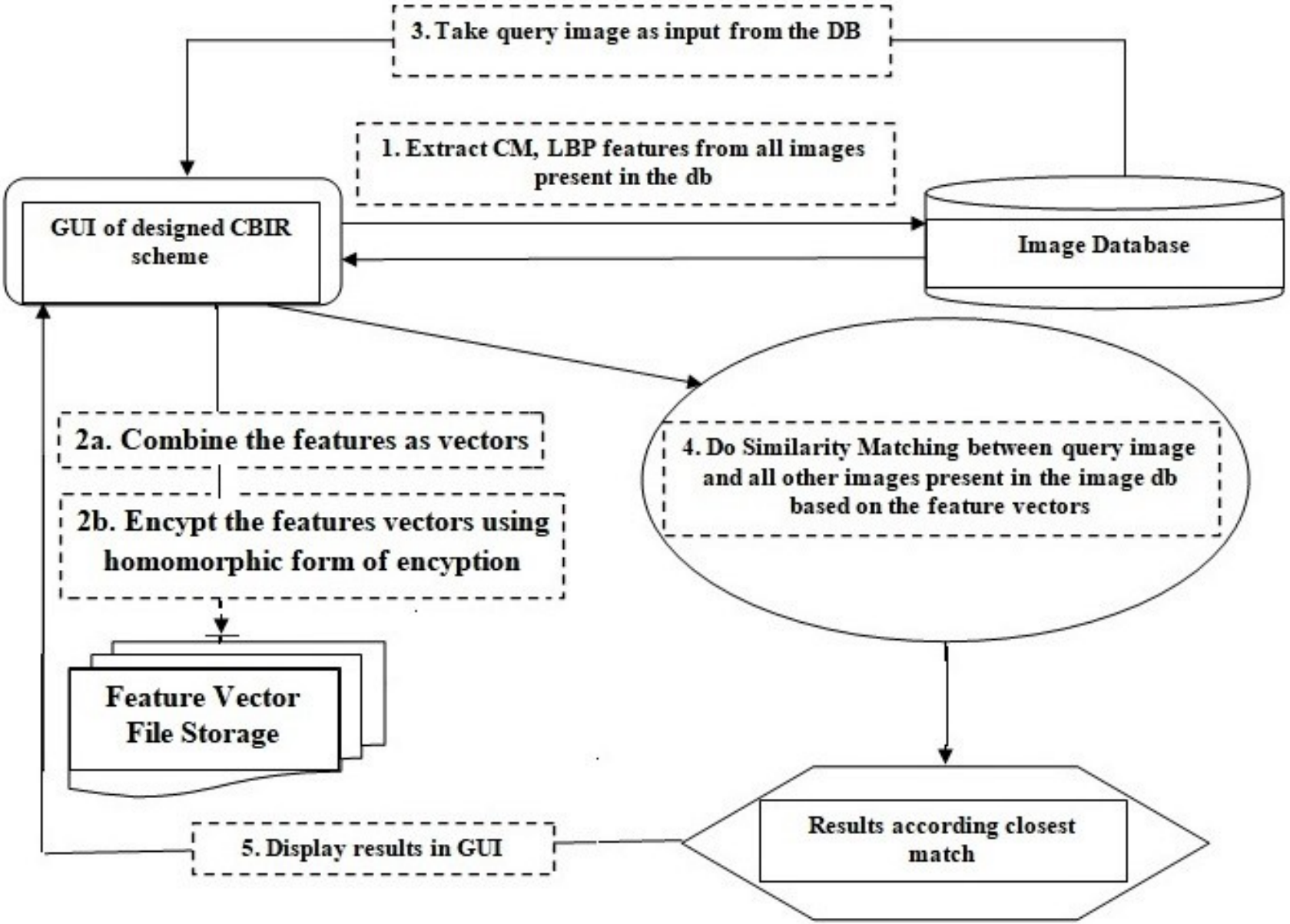
The Problem Statement

- **Problem** - Encryption causes CBIR schemes in non-encrypted domain to become unusable.
- **Proposed Solution** - A scheme that supports CBIR using the encrypted features of an image.
  a. Extracts both the color (Color moment) & texture (local binary pattern) features from images & combines them as a feature vector.
  b. Feature vectors are protected by using a homomorphic form of encryption before being stored.
  c. The similarity matching between query image & stored images should be done in such a way that, it should not be affected by encryption.
- The plan is to use a modified version of hamming distance for this cause.

In Theory

- There are three main sub-parts in this scheme: 1. Feature extraction, 2. Feature Encryption and 3. Searching/Retrieval.
- **Feature Extraction** - CM and LBP are extracted for each image in image database.
- **Feature Encryption** - Encrypt CM & LBP using some homomorphic form of encryption & store the encrypted feature vectors in a file storage.
- **Searching** - Initially the same 'Feature extraction' algorithm is employed in the query image, then the features are encrypted using same homomorphic form of encryption & subsequently the similarity is measured with the feature vector storage using appropriate similarity measure.

Process Flow Diagram of proposed scheme



Color Moments

- Color moments calculate distribution of color in an image.
- Reusing method of Stricker and Orengo :
- If value of a^{th} color channel, at b^{th} image pixel is P_{ab},
- The first order color moments or **mean** is:

$$E_a = \frac{1}{N}\sum_{b=1}^{N} P_{ab}$$

- The second order color moments or **standard deviation** is:

$$\sigma_a = \sqrt{(\frac{1}{N}\sum_{b=1}^{N} P_{ab} - E_a)^2)}$$

- The third order color moments or **skewness** is:

$$S_a = \sqrt[3]{(\frac{1}{N}\sum_{b=1}^{N} (P_{ab} - E_b)^3)}$$

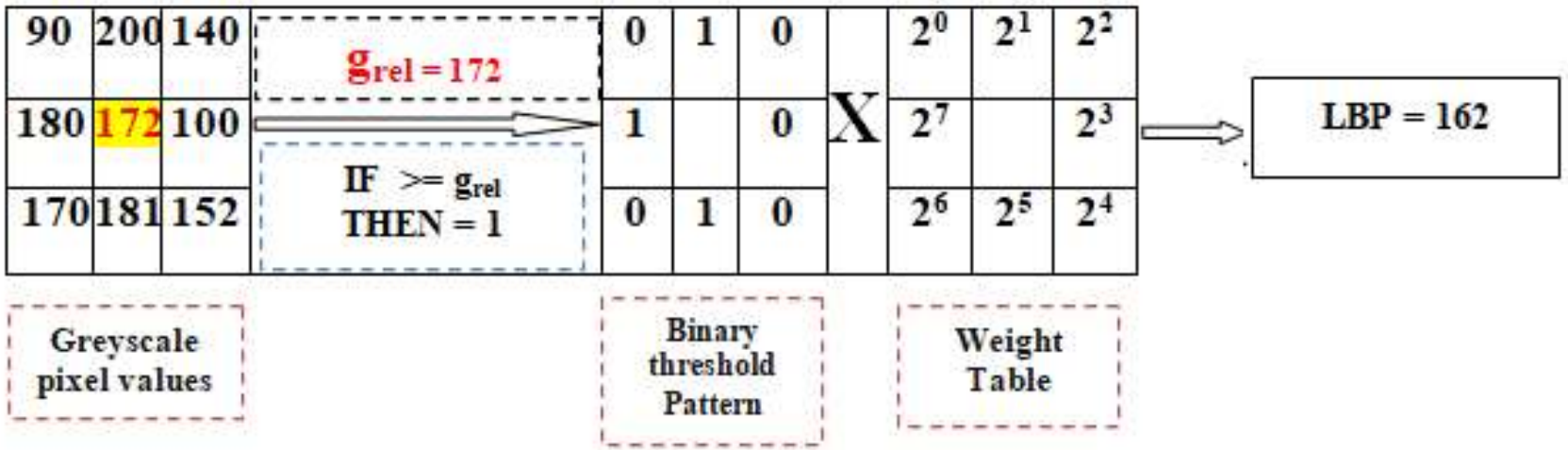
Local Binary Pattern (LBP)

- Introduced by Ojala et al [5], created setting a threshold.
- Illustration : P = 7 and p = 0, 1, 2, 3, 4, 5, 6, 7

$$LBP = \sum_{p=0}^{P-1} s(g_p - g_{rel}).2^p \quad \text{And,} \quad s(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

for

90	200	140
180	172	100
170	181	152



Here g_p in each block is thresholded by its center pixel value g_{rel}. And S(x) is the Binary threshold pattern function.

Achieved So Far...

Algorithm 1 : ExtractFeatures

Input: Database of Images.
Output: A file storage filled with extracted feature vectors.

Begin

For all images in the database provided -

a. Calculate the color moments (cm).
b. Convert the image into a gray-scale image (if the image is a colored one).
c. Calculate the local binary pattern (lbp) in vectorized form.
d. Calculate the feature vector as f = [cm lbp]

End For

Save the feature vectors in a file storage.

End

Algorithm 2 : RetriveMatchingImages

Input: Query Image.
Output: Retrieved images closest to the queried one.

Begin

Extract feature vector for the query image.

For all images in the database provided

a. Calculate the similarity among the image and query image

End For

Pick a certain pre-fixed number of images from the image database closest to the Query image based on the used similarity measure.

Show the picked images in the retrieval UI.

End

- Tested with 1000 image samples among 10 classes - humans(Africans), beaches, monument buildings, dinosaurs, buses, elephants, flowers, horses, mountains & foods.
- One of sample result of the algorithm for closest 10, 15 and 20 matches respectively :

Figure : – Closest 10 Matches

Figure : – Closest 15 Matches

Figure: – Closest 20 Matches

- Perform quantitative analysis, using precision and recall values defined as :
- Precision = $\frac{No.of\,Relevant images retrieved}{Total No.of images retrieved} \times 100$
- Recall = $\frac{No.of\,Relevant images retrieved}{No.of relevant images in the database} \times 100$
- The Calculated precision & recall values in a tabular form are : -

Type of Images	Precision Values for retrieved images for			Recall Values for retrieved images		
	10 matches	15 matches	20 matches	10 matches	15 matches	20 matches
Humans	0.8	0.87	18/20	0.08	0.13	0.18
Beaches	0.6	0.6	0.45	0.06	0.09	0.09
Monuments	0.5	0.47	0.5	0.05	0.07	0.1
Buses	0.5	0.47	0.5	0.05	0.07	0.1
Dinosaurs	1	1	1	0.1	0.15	0.2
Elephants	0.1	0.2	0.15	0.01	0.03	0.03
Flowers	0.8	0.8	0.85	0.08	0.12	0.17
Horses	0.9	0.9	0.85	0.09	0.14	0.17
Mountains	0.3	0.4	0.4	0.03	0.06	0.08
Foods	0.3	0.3	0.4	0.03	0.05	0.08

Table -1: Precision & Recall Values of Tested Data Sets

Future Scope

- Enhance the CBIR scheme, so that in can be used for encrypted images also that uses an homomorphic form of encryption.
- Rather than calculating color moments from the whole image, extract this value on a block level from each image.
- Use a modified version of hamming distance rather than currently used euclidean one to have better similarity matching.
- Make sure that the used distance measure is applicable to images that has gone through homomorphic form of encryption.
- Security Analysis on the enhanced CBIR scheme.
- Performance evaluation for the encrypted images.

References

[1] Annesha Malakar, Joydeep Mukherjee, ”Image Clustering using Color Moments, Histogram, Edge and K-means clustering” , *International Journal of Science and Research , Vol.2, No. 1*, pp. 532-537, Jan. 2013.

[2] Aman Chadha , Sushmit Mallik, Ravdeep Johar, ”Comparative Study and Optimization of Feature-Extraction Techniques for Content based Image Retrieval” , *international Journal of Computer Applications*, Vo152, No .20, August 2012.

[3] Zhi-chun huang, Patrick P. K. Chan, Wing W. Y. Ng, Daniel s. Yeung ”Content-based image retrieval using color moment and Gabor texture feature”, *Proceedings of the Ninth international Conference on Machine Learning and Cybernetics (IEEE)*, Vol. 2, pp 719-724, July 2010.

[4] Caifeng Shan, ”Learning local binary patterns for gender classification on real-world face images”, *Pattern Recognition Letters*, Vo1.33, No.4, pp 431-437, March 2012.

[5] T. Ojala, M. Pietikinen, D. Harwood, ”A comparative study of texture measures with classification based on featured distributions”, *Pattern Recognition*, Vo1.29, No. 1, pp. 51-59, 1996.

THANK YOU