A COMPACT IMAGE SIGNATURE FOR RTS-INVARIANT IMAGE RETRIEVAL

Pratim Ghosh*, B.S.Manjunath[$], K.R.Ramakrishnan*

*Indian Institute of Science, Bangalore, India, (prathim,krr@ee.iisc.ernet.in) ,
[$]University of California Santa Barbara, USA, (manj@ece.ucsb.edu).

Keywords: Fourier-Mellin Transform, principal component analysis, Lloyd max scalar quantization, RTS invariant compact image signature, duplicate image retrieval.

Abstract

In this paper we propose a novel framework to obtain a very compact image signature (32 bits) which is invariant to rotation, translation, scaling and other minor perturbations like smoothing, random noise addition, JPEG compression etc. The framework involves Fourier-Mellin transform, conventional PCA and non-uniform scalar quantization. The high retrieval efficiency and low space consumption demonstrates the significance of our signature in duplicate image retrieval and large image database indexing.

1 Introduction

Rapid growth in multimedia and networking technology has created an urgent need for developing a system to mitigate copyright infringement. Thanks to the advances in electronic media technology one can easily copy another's creativity and distribute it over the internet in his name. This kind of infringement is more rampant for image and video.

Existing duplicate detection systems are of two types: watermarking based [4] and retrieval based [7, 8]. In the former some data is hidden with in the image and in such a scenario the emphasis is on finding a proper trade off between distortion and robustness. But in retrieval based framework the emphasis is on evaluating good feature representation of images. Duplicate detection involves retrieving either exactly identical images or near identical (slightly perturbed (rotated, scaled, translated)) version of given reference image from a large database. In this work we also consider two images to be nearly identical if one is obtained by subjecting the other to slight smoothing, blurring and cropping kind of operations.

In order to identify a duplicate this paper attempts to find a good compact descriptor (referred to as signature after quantization), which is invariant to all the operations mentioned above. Several papers [3, 6] have talked about local descriptors but it has been found difficult to obtain a compact descriptor using local approaches. The number of interest points can be as high as thousand or so depending on the complexity of the images. Also local descriptors can give false alarm if the database contains images of the same scene taken from different view directions and images with several similar regions. Handling global descriptors has also been exhaustively discussed in several works [7, 8]. However the compaction of these descriptors seems to have remained unaddressed. This paper proposes an approach towards this end.

The remainder of the paper is organized as follows. Some background information and proposed framework are explained in Section 2. Section 3 presents the experimental results and discussion. Finally, our conclusion and ideas for future work are presented in Section 4.

2 RTS invariant signature

The overall block diagram of the image signature extraction is depicted in Figure 1.

2.1 RTS invariant feature extraction

Fourier-Mellin transform (FMT) has been studied extensively by pattern recognition community. It has also been used a lot in watermarking and several other applications. Few years ago, the problem related to computation of FMT was solved by using Analytical Fourier-Mellin Transform (AFMT) [1]. We have used fast AFMT approximation (F-AFMT) [2] for our work in this paper.

Let $f_1(x, y)$ be an image and its rotated, translated and scaled version $f_2(x, y)$ be related by the equation

$$f_2(x,y) = f_1(\alpha(x\cos\beta+y\sin\beta)-x_t, \alpha(-x\sin\beta+y\cos\beta)-y_t). \quad (1)$$

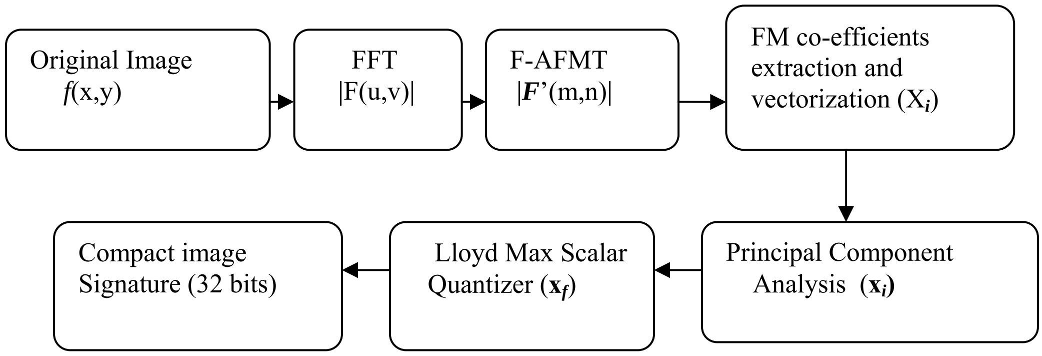


Figure 1: Overview of proposed framework.

where the RTS parameters are β, (x_t, y_t), α respectively. Applying Fourier transform (FFT) on both sides of (1) and taking only the magnitude we get

$$|F_2(u,v)|=|\alpha|^{-2}|F_1(\alpha^{-1}(u\cos\beta+v\sin\beta),\alpha^{-1}(-u\sin\beta+v\cos\beta))|. \quad (2)$$

where (u, v) correspond to variables in frequency domain. Now taking F-AFMT on both sides of (2) is equivalent to log-polar transform followed by another 2D Fourier transform on $|F_2(u, v)|$. Substituting

$$u=e^{\rho}\cos\psi. \quad (3)$$

$$v=e^{\rho}\sin\psi. \quad (4)$$

in (2) we get

$$|F_2(u,v)|=|\alpha|^{-2}|F_1(\alpha^{-1}e^{\rho}\cos(\psi-\beta),\alpha^{-1}e^{\rho}\sin(\psi-\beta))|. \quad (5)$$

$$=|\alpha|^{-2}|F_1(e^{(\rho-\log\alpha)}\cos(\psi-\beta),e^{(\rho-\log\alpha)}\sin(\psi-\beta))|. \quad (6)$$

or

$$|F_2(\rho,\psi)|=|\alpha|^{-2}|F_1(\rho-\log\alpha,\psi-\beta)|. \quad (7)$$

The above expression demonstrates that the amplitude of the log-polar spectrum is scaled by a constant factor (can be handled by normalizing the co-efficient) $|\alpha|^{-2}$ and that image scaling and rotation in spatial domain result in translation by $\log\alpha$ and β in log polar domain (Figure 2). In order to get rid of the translation parameters on the right hand side of (7) we apply Fourier transform again to (7) and taking the magnitude we get

$$|\boldsymbol{F}'_2(m,n)|=|\alpha|^{-2}|\boldsymbol{F}'_1(m,n)|. \quad (8)$$

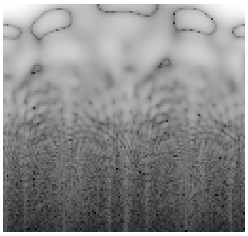
where $\boldsymbol{F}'_1$ and $\boldsymbol{F}'_2$ are the FFT of F_1 and F_2.

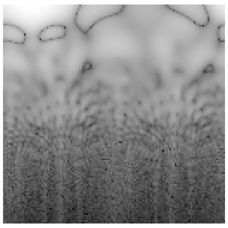
In view of the above, to obtain an RTS invariant set of features we apply on a given image FFT and F-AFMT successively. We then extract all Fourier Mellin (FM) co-efficient (except the D.C component) lying with in a fixed radius from the D.C component. After several experiments the radius was fixed at a value, which encompasses around 50% of the total A.C energy. It was found that for all practical purposes this is sufficient to represent an image (Figure 3). The value of the radius chosen corresponds to a good trade off between initial compaction and retrieval accuracy shown later. A 64*64 grid has been found adequate for log-polar mapping. Let X_i (m dimensional) denote the vector of extracted FM co-efficient corresponds to i^{th} image (Figure. 1).

We now apply PCA [10] to the set of vectors $X=[X_1, X_2, \ldots, X_N]$ where N (<< size of database) is set of training images selected randomly from the database. After applying PCA we get new feature vectors $\mathbf{x}_i$ ($i=1, 2, \ldots, N$) of length l ($l \ll m$).

The $\mathbf{x}_i$ obtained above is subjected to scalar quantization. A Llyod Max quantizer [5] is designed for each component of $\mathbf{x}_i$ vectors. The signature of the i^{th} image is obtained by concatenating the l quantized components of the $\mathbf{x}_i$ vector. Choosing a value of 4 for l and 8 bit quantizers for the four components yields a signature of length 32 bits.

(a) **(b)**





(c) **(d)**

Figure 2: (a) Original image (b) Rotated and Scaled Image (c)&(d) log-polar map of Fourier spectrum of (a)&(b) respectively. Results showing rotation and scaling in spatial domain corresponds to translation in log-polar domain.

3 Experimental results and discussion

In this section we first describe the evaluation metric used to asses the performance of our signature. Then we proceed to present experimental results.

3.1 Evaluation metric

Precision-recall curve has been used to measure the performance of our signature. Let $\mathbf{A(T,\Gamma)}$ be the set of $\mathbf{T}$ retrievals based on the smallest distances from $\mathbf{\Gamma}$ (query image) in the signature space and $\mathbf{R(\Gamma)}$ is the set of $\mathbf{D}$ images in the database relevant to $\mathbf{\Gamma}$. Then, *precision* which is defined by $\mathbf{P}$ is the proportion of images retrieved that are relevant to query image $\mathbf{\Gamma}$.

$$\mathbf{P(T,\Gamma)} \underline{\underline{\Delta}} \frac{|\mathbf{A(T,\Gamma)} \cap \mathbf{R(\Gamma)}|}{\mathbf{T}}$$

Recall which is defined as

$$\mathbf{C(T,\Gamma)} \underline{\underline{\Delta}} \frac{|\mathbf{A(T,\Gamma)} \cap \mathbf{R(\Gamma)}|}{\mathbf{D}}$$

is the proportion of relevant images retrieved from $\mathbf{R(\Gamma)}$. *Precision-recall* curve is plotted by averaging *precision* over all $\mathbf{\Gamma}$. *Precision-recall* curves for different cases are shown in Figure 4 and Figure 5. In each case average *precision* values are plotted at *recall* values $\frac{1}{\mathbf{D'}}, \frac{2}{\mathbf{D'}}, \ldots\ldots, \frac{\mathbf{D}}{\mathbf{D'}}, \frac{\mathbf{D+1}}{\mathbf{D'}}, \ldots\ldots \mathit{upto}\ 1$ (where $\frac{\mathbf{D}}{\mathbf{D'}}$ is defined as knee point and $\mathbf{D} < \mathbf{D'}(\text{integer}) \leq \mathbf{2D}$). Here *recall* is treated as independent variable whereas *precision* is dependent one.

3.2 Results

To study the performance of the proposed scheme we used the database MM270k. This database can be downloaded from http://www-2.cs.cmu.edu/~yke/retrieval. It contains 18,785 images including landscapes, animals, construction and people. The aspect ratio and image sizes vary over the database. The images are mostly in colour except for a few gray ones.

For our experiments we created a database containing around 5000 images (randomly chosen from MM270k).

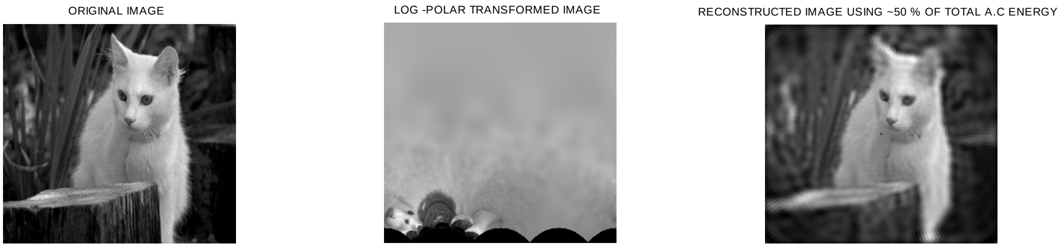


Figure 3: These are original image, log-polar transformed image and reconstructed image (from left to right) using only ~50 % of total A.C energy. Overall shape remains unchanged in the reconstructed image.

We chose 100 random images (queries) and for each of these query images 13 duplicates were generated by performing all the operations mentioned in Table 1.

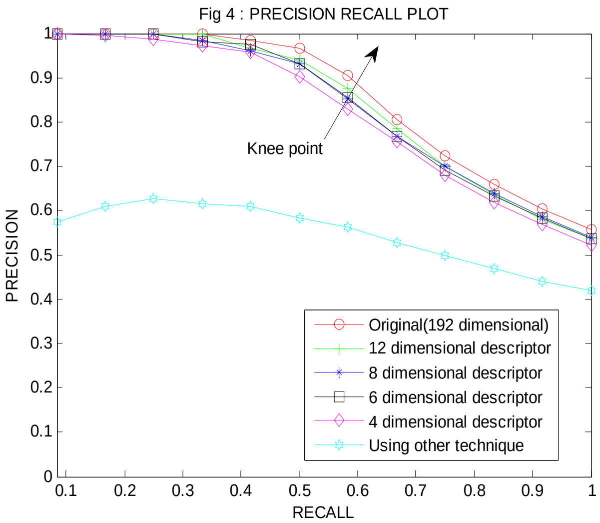


Figure 4: *Precision-Recall* plot.

For data hiding we used Image Adaptive QIM (Quantized Index Modulation) technique. The original descriptor for each image was 192 dimensional. We have plotted the *precision-recall* curves for the first seven operations mentioned in Table 1. The plots are obtained by averaging over the 100 query images. Figure 4 shows the results for original descriptor and reduced feature descriptors (using PCA). These include 12, 8, 6 and 4 dimensional descriptors. Figure 5 shows the results after applying non-uniform and uniform scalar quantization on the reduced feature descriptors. Each quantized feature in the descriptor was represented by 8 bits. (Thus for a 4 dimensional descriptors a 32 bits signature is obtained) *Precision* values are above 80% till the knee point and droop afterwards in both of the Figure 4 and Figure 5. So plots essentially capture the high retrieval performance of our signature. Table 1 shows the variation of compact image signature with respect to the original one after various operations. The value of **D'** chosen for the plots is 12. However other values of **D'** have also shown to yield similar results.

To compare our performance we chose another RTS invariant feature descriptor [9] based on scale normalization of power spectral density by a cut off frequency followed by Zernike moments calculation. This method gives an image descriptor of length 22 (8x22=176 bits). We have used **L1- Norm** for distance calculation. The whole algorithm is implemented in Matlab. It takes approximately 0.4 seconds to extract the signature on Pentium –III, 1 GHz machine.

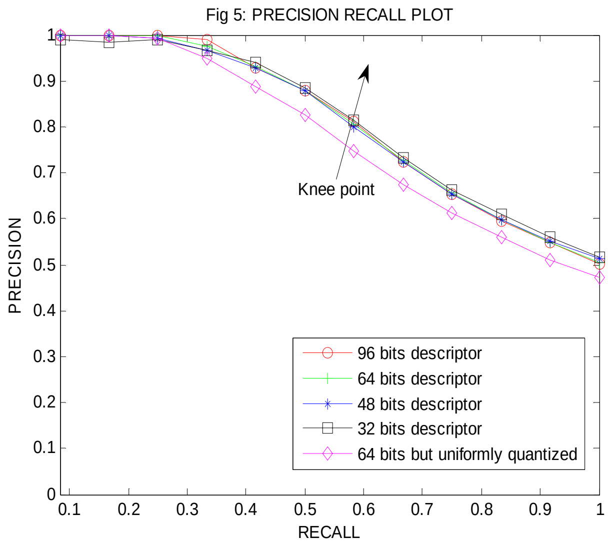


Figure 5: *Precision-Recall* plot.

4. Conclusion

In this paper a compact signature for an image has been defined and has been shown to give good retrieval accuracy from a database. The signature is insensitive to various operations like rotation, blurring, scaling, random noise addition and JPEG compression. However when one uses the signature on a stegoed duplicate (Table 1), the signature vary to the extent it appears possible to detect stegoed duplicates.

No.	Operations Performed	Final Image Signature
0.	Original Image	186 81 167 154
1.	Gaussian Noise	186 81 166 154
2.	Rotation (180)	186 82 164 156
3.	Rotation (270)	186 82 164 156
4.	Gaussian Blurring(2)	186 80 167 153
5.	Scaling down (x1.3)	185 79 166 154
6.	Scaling down (x2)	183 79 163 154
7.	Scaling down (x4)	180 86 156 153
8.	Hiding Data (1 KB)	179 93 166 165
9.	Hiding Data (2 KB)	179 93 165 166
10.	JPEG Compressed(30)	186 82 167 155
11.	JPEG Compressed(50)	186 81 167 156
12.	JPEG Compressed(70)	186 81 167 154
13.	JPEG Compressed(90)	185 81 167 155

Table 1: Comparison of compact image signature of modified images with respect to the original one for a particular category.

The image signature proposed here does have certain handicaps. The signature is tied to a database in the sense that the statistics of the database play a crucial role in defining the signature. It may not be useful when applied to a totally new database. Efforts are on to define a universal signature, which can work satisfactorily on any database.

References

[1] D. Casasent and D. Psaltis, "Scale invariant optical transform", Opt.Eng. Vol-15, no-3, pp. 258-261, (1976).

[2].S. Derrode and F. Ghorbel, "Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description." *CVIU*, Vol. 83, no-1, July, (2001).

[3]Y. Ke and R. Suthankar, "Efficient Near Duplicate Detection and Sub Image Retrieval", ACM Multimedia Conference, August (2004).

[4] Ching-Yung Lin, Min Yu, Jeffrey A. Bloom, Ingemar J. Cox, Matt L. Miller and Yui Man Lui, "Rotation Scale and Translation resilient Watermarking for images" , IEEE Transaction on Image Processing, Vol-10, no 5, pp. 767-782, May (2001).

[5] Stuart P.Lloyd, "Least Squares Quantization in PCM", IEEE Transactions on Information Theory, Vol-28, no-2, pp.129-137, March (1982).

[6] D. Lowe. "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, Vol. 60, no-2, pp. 91–110, (2004).

[7]Y. Maret, F. Dufaux, and T. Ebrahimi, "Image Replica Detection based on Support Vector Classifier," Optical Information System III, Proceedings of SPIE, Vol-5909 pp. 173-181, (2005).

[8] Sujoy Roy, Ee-Chien Chang, "A unified framework for resolving ambiguity in Copy detection", ACM Multimedia Conference, pp. 648-655, (2005).

[9] Dong-Gyu Sim, Hae-Kwang Kim and Dae-I*l* Oh, "Translation, Scale and rotation invariant texture descriptor for texture based image retrieval," 2000 International Conf. on Image Processing, 2000 Proceedings, vol-3, pp. 742-745.

[10] Matthew Turk and Alex Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuro Science. Vol –3, no-1, pp. 71-86, (1991).