LOW COMPLEXITY RST INVARIANT IMAGE RECOGNITION USING FOURIER MELLIN TRANSFORM

P. Ayyalasomayajula, S. Grassi, and P.-A. Farine

Institute of Microengineering, Electronics and Signal Processing Laboratory, Ecole Polytechnique Fédérale de Lausanne,
EPFL STI IMT ESPLAB, Rue A.-L. Breguet 2, 2000, Neuchâtel, Switzerland.
phone: + (41) 32 718 3425, fax: + (41) 32 718 3402, email: pradyumna.ayyalasomayajula@epfl.ch
web: http://esplab.epfl.ch

ABSTRACT

In this paper we propose a low complexity method for Rotation, Scale and Translation (RST) invariant content-based image retrieval, suitable for a handheld image recognition device. The RST compensation method is based on Fourier-Mellin Transform (FMT) which we implement efficiently using log-polar grid interpolation. This RST compensation method is used in conjunction with an image recognition algorithm based on Discrete Cosine Transform (DCT) phase matching. A pre-selection algorithm is also added for decreasing the complexity. This algorithm is based on color proportions within concentric circular zones encompassing the edge pixels. The resulting RST invariant image recognition system was tested on 1500 pictograms and 1000 pictures with different RST conditions, showing an average recognition accuracy of 95.2% for pictograms and 96.9% for pictures.

1. INTRODUCTION

As the number and the size of digital image databases are increasing, the development of image retrieval systems is becoming increasingly important. Among them, Content-Based Image Retrieval (CBIR) has gained lot of interest from researchers starting from the early 1990's [1]. CBIR uses information ("features") extracted from the image, to describe the image and to search for its closest match within a database of possible images.

CBIR can be used for handheld image recognition devices in which the image to be recognized ("query") is acquired with a camera, and thus there is no additional metadata associated to it. However, due to the camera acquisition, the query image is subject to geometric transformations, such as Rotation, Scaling and Translation (RST) with respect to the reference version stored in the database, thus the need to develop a CBIR system which is invariant to RST transformations. An additional requirement of handheld image recognition devices is low complexity, in order to have low power consumption as well as a fast response when searching large databases. Usually, the features used for CBIR are not robust to RST transformations, resulting in a wrong match if they are used directly for the search.

Different methods such as histogram descriptor [2] and Scale Invariant Feature Extraction [3] are proposed to achieve RST feature invariance. But these methods need numerous computations and are not suitable for handheld recognition applications. Fourier-Mellin Transform (FMT) for image recognition was initially reported by Y. Sheng et al. [4] and Chen et al. [5] and was used to extract RST invariant features. In recent years we have seen an increased use of FMT for extracting RST invariant features [6], [7].

In this paper we extend the Fourier-Mellin based methods and propose a low-complexity RST invariant image recognition system using Fourier-Mellin Transform and DCT phase matching [8] along with a pre-selection stage, based on color proportions within circular concentric zones encompassing the edge pixels [9]. We also propose an efficient implementation of FMT using log-polar grid interpolation for handheld recognition devices.

The paper is organized as follows. The proposed system for image recognition is described in Section 2. Section 3 briefly explains the Fourier-Mellin Transform. Section 4 presents the efficient implementation of Fourier-Mellin Transform using log-polar grid interpolation. Section 5 reports on the experimental evaluation of the proposed image recognition system. Conclusions are drawn in Section 6.

2. PROPOSED IMAGE RECOGNITION SYSTEM

Figure 1 illustrates the proposed image recognition system. This system can be divided into three stages. In the first stage, to decrease the complexity, we use for pre-selection an RST invariant, less complex but also less accurate algorithm called "Color Density Circular Crop" (CDCC). The CDCC algorithm (see *Subsection 2.1*) searches the database and preselects the 50 images which are closest to the query image. In the second stage the query image is RST compensated with respect to each of the images of the reduced database containing the 50 pre-selected images coming from CDCC. This stage involves extraction of RST parameters using Fourier-Mellin Transform and is explained in Sections 3 and 4. In the third stage, a more accurate but computationally more expensive algorithm called "DCT Phase Match" (DCTPM) is used to improve the overall recognition accuracy of the system. The DCTPM algorithm (see *Subsection 2.2*) performs correlation between the RST compensated query image and each of the images of the pre-selected database, to find the best match.
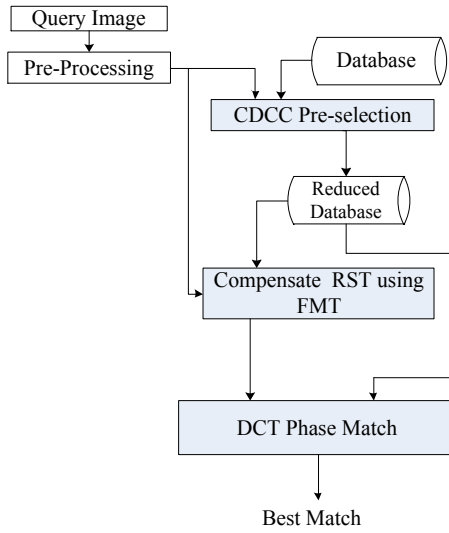
Figure 1 – Overview of the proposed image recognition system

Figure 2 - An example of circular segmentation and concentric cropping of an image into 4 concentric regions.

Figure 3 – DCT phase match

2.1 Color Density Circular Crop (CDCC) pre-selection

The CDCC pre-selection algorithm [9] searches the database and preselects the 50 images which are closest to the query image. CDCC uses as feature the color proportions within concentric circular zones of the image. To compute the CDCC features of an image we first perform canny edge detection. Once the edge pixels in an image are found, the center of a rectangle, $\{x_c, y_c\}$, which is formed by using the extrema edge pixels in both axes is calculated as shown in Eq. 1:

$$\{x_c, y_c\} = \left\{ \frac{x_{Max} + x_{Min}}{2}, \frac{y_{Max} + y_{Min}}{2} \right\} \qquad (1)$$

Here, x_{Max} and x_{Min} are the extrema points along the X-axis of the image, and y_{Max} and y_{Min} are the extrema points along the Y-axis.

The distance from this center to the farthest edge pixel is then computed. This distance and the center $\{x_c, y_c\}$ define a circle which tightly encompasses all the edge pixels. This circle is then divided into four concentric regions and the color density of the red, green and blue channels for each concentric region are calculated. These color densities are assembled in a feature vector of length $L = 12$ which uniquely represents the image. *Figure 2* shows an example of circular segmentation and concentric cropping of an image.

For every image stored in the database, the feature vector is pre-calculated and stored. When a query image is acquired, the feature vector of the query is calculated and compared using L^1 distance with the feature vector of each of the images stored in the database. As the feature vector length is small and also low in complexity to 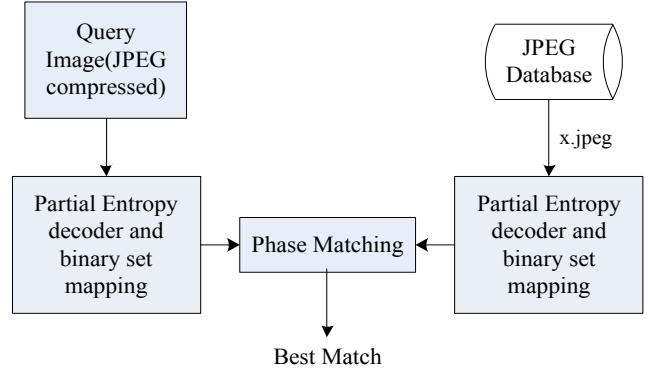compute, the CDCC algorithm is low in complexity. Additionally, this algorithm is robust to rotation, translation and scaling, but is sensitive to variations in lighting.

2.2 DCT Phase Match (DCTPM)

The base recognition algorithm, which is referred to as DCT phase match [8], searches for the best match of the query image into a database using correlation on the DCT phase of the 8x8 blocks of the images, and thus is compatible with the JPEG compression standard. This algorithm is accurate and robust to lighting variation but computationally expensive. To reduce complexity, DCT phase match is only performed on the reduced database of the 50 images, preselected by the CDCC algorithm, with the RST compensated query image. *Figure 3* shows the main blocks of the DCT phase match method.

3. FOURIER-MELLIN TRANSFORM (FMT)

Fourier-Mellin Transform is used to extract the RST parameters which are then used to compensate the query image before comparing it (using DCT phase matching) with each of the 50 candidate images of the pre-selected database. The RST parameters are the scale parameter (s), the rotation angle (θ) and the translation parameter (Δx, Δy). In order to calculate them, consider a two-dimensional reference grey-scale image f and the RST transformed image f_{RST} which are related by the transformation $f_{RST} = T(f)$. The relation of each pixel $f(x, y)$ of f which maps to a corresponding pixel $f_{RST}(x_t, y_t)$ of f_{RST} is given by:
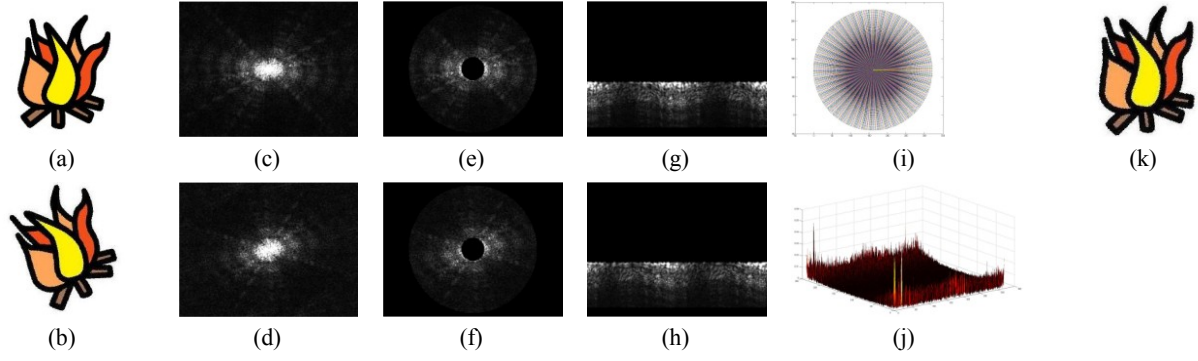
Figure 5 – (a) Reference image; (b) Rotated and scaled image; (c),(d) Center-shifted Fourier transform of (a) and (b) respectively; (e),(f) Band-pass filtered version of (c) and (d); (g),(h) Log-polar mapping of (e) and (f); (i) Log-polar mapping grid; (j) Phase correlation between (g) and (h); (k) RST compensated image.

$$\begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix} = \begin{bmatrix} s\,cos\theta & -s\,sin\theta & \Delta x \\ s\,sin\theta & s\,cos\theta & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

From Eq.2, we have:

$$f_{RST}(x_t, y_t) = f(s \cdot (x \cdot cos\theta - y \cdot sin\theta) + \Delta x, \\ s \cdot (x \cdot sin\theta + y \cdot cos\theta) + \Delta y) \quad (3)$$

$$\mathcal{F}_{RST}(u, v) = \frac{e^{-j2\pi(\Delta x \cdot u + \Delta y \cdot v)}}{s^2} \cdot \\ \mathcal{F}\left(\frac{u \cdot cos\theta - v \cdot sin\theta}{s}, \frac{u \cdot sin\theta + v \cdot cos\theta}{s}\right) \quad (4)$$

where $\mathcal{F}$ and $\mathcal{F}_{RST}$ are the Fourier Transform of f and f_{RST}, respectively. Mapping to log-polar domain using $u = e^\rho cos\varphi$ and $v = e^\rho sin\varphi$, and taking the magnitude of both sides of the equation, we have, from the translation and scale property of Fourier transform:

$$|\mathcal{F}_{RST}(e^\rho cos\varphi, e^\rho sin\varphi)| \\ = \frac{1}{s^2}\left|\mathcal{F}\left(e^{\rho - log\,s}cos(\theta + \varphi), e^{\rho - log\,s}sin(\theta + \varphi)\right)\right| \quad (5)$$

$$\mathcal{F}_{RST-LP}(\rho, \varphi) = \frac{1}{s^2}\mathcal{F}_{LP}(\rho - log\,s, \theta + \varphi) \quad (6)$$

Where $\mathcal{F}_{RST-LP}$ and $\mathcal{F}_{LP}$ are the log-polar mapping of $|\mathcal{F}_{RST}|$ and $|\mathcal{F}|$ respectively.

From Eq. 6 we can see that the rotation (by θ) and the scaling (by s) are now translations in the mapped domain, and we can measure this rotation (θ) and scale (s) by performing a 2D phase correlation [10]. After the rotation and scale parameters are extracted we can compensate the image for these extracted parameters and then run another 2D phase correlation to extract the translation parameter (Δ x, Δ y).

4. IMPLEMENTATION OF FMT

To efficiently implement Fourier-Mellin transform for extracting the RST compensation parameters, we implement the log-polar mapping seen in the previous section by using log-polar grid interpolation (see *Subsection 4.1*).

The algorithm for RST extraction using FMT is as follows:

1. Calculate the Fourier transform of the query image and the candidate reference image from the database and shift the zero-frequency component to the center of the spectrum.
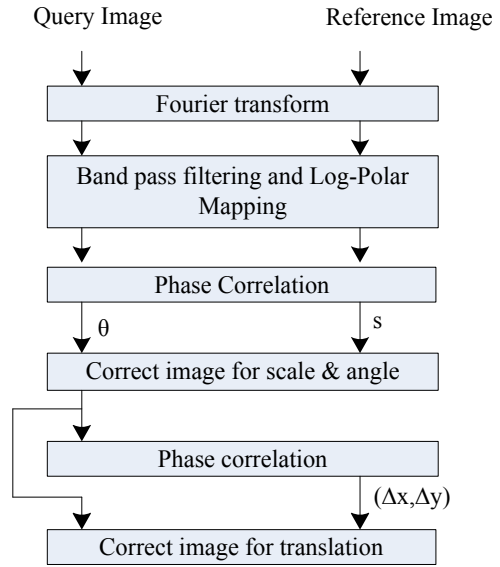

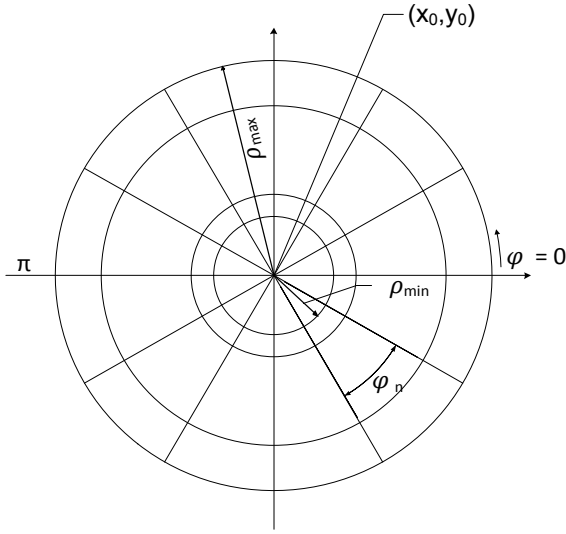Figure 4 – Fourier-Mellin transform implementation for RST compensation

Figure 6 – Log-polar mapping grid with logarithmically spaced radius (ρ) and linearly spaced angles (φ) with center (x$_0$,y$_0$)

2. Perform band-pass filtering with a minimum radius of $\rho_{\min}$ and maximum radius $\rho_{\max.}$, and then log-polar mapping of the magnitude of the Fourier transform. The radii ρ_{min} and ρ_{max} for the band-pass filter are calculated as 0.04 and 0.8 times of the minimum distance from the center of the image to the corner of the image.

3. Perform phase correlation between the two log-polar mapped images, and extract the rotation (θ) and scale (s) parameters.

4. Compensate the query image for rotation and scaling, using the extracted rotation and scale parameters.

5. Perform phase correlation on the compensated query image with respect to the reference image to extract the translation parameters (Δ x, Δ y).

Figure 4 shows the flowchart for the implementation and *Figure 5* shows an example of the RST parameters extraction and compensation using FMT.

4.1 Log-polar grid interpolation

The log-polar grid is obtained by quantizing the two parameters of Eq. 6, which are the radius, ρ, and the angle, φ. The radius is quantized by dividing logarithmically the distance from the center of the image (x$_0$, y$_0$) to the corner of the image. The angle is linearly quantized between 0 and 2π. That is, in Matlab commands:

ρ = logspace(0, log(d))

φ = linspace(0, 2π)

The amount of points in the log-polar grid is chosen accordingly to the resolution of the image, so as to have the same amount of points as in the sampling grid of the frequencies of the image ((u, v) of Eq. 4).

By using FFT, we have already calculated the value of the transforms at all the points of the (u, v) sampling grid. In order to calculate the values of the transforms at the points of the (ρ, φ) log-polar grid we proceed as follows. We first map the points of the (u, v) sampling grid into points in the (ρ, φ) log-polar domain. We then know the value of the transforms at these points, but these points are not necessary on the (ρ, φ) log-polar grid. We thus obtain the values for each point of the log-polar grid by linear interpolation using the known values of neighboring points of the mapped (u, v) grid.

Figure 6 shows an example of a log-polar grid which has been filtered with a band pass filter with a minimum radius of $\rho_{\min}$ and maximum radius $\rho_{\max}$. The grid points are at the intersection of the radial lines with the concentric circles. The actual log-polar grid we have used is shown in *Figure 5i*.

5. EXPERIMENTAL EVALUATION

In the next sub-sections we explain the databases used for experimental evaluation, testing procedure and results.

5.1 Experimental Databases

Two databases, one for pictograms and one for pictures were used during the tests. The pictogram database contains 1500 pictograms from the Picture Communication Symbols set [11]. The pictograms were stored in JPEG format with 85% quality, and QVGA resolution (320 x 240). The picture database contains the 1000 pictures from the COREL photograph data set used in [12] which are also JPEG compressed color images with QVGA resolution. We refer to these two databases as the "clean pictogram database" and the "clean picture database".

Additionally, a set of 50 representative pictograms and 50 representative pictures were chosen from the clean databases. These images were printed and acquired using a fixed focus OV7675 VGA camera from Omnivision. The images were manually aligned and cropped using a printed visual reference [13]. These two sets of 50 images constitute what we call the "real condition pictogram database" and the "real condition picture database".

		Pictograms	Pictures
Clean database	No RST	100.0 %	100.0 %
	Rotation $\theta = 30°$	93.9 %	96.8 %
	Rotation $\theta = 60°$	94.6 %	97.2 %
	Rotation $\theta = 90°$	99.6 %	99.8 %
	Scale s = 0.5	88.5 %	90.7 %
	Scale s = 0.75	91.1 %	95.4 %
Translation($\Delta x, \Delta y$) = (100,100)		99.1 %	98.9 %
Real condition database		90.0 %	94.0 %

Table 1 – Summary of results from the image recognition system for 'real condition database' and under different RST conditions on 'clean database'

772

5.2 Testing Procedure

The proposed image recognition system was implemented in Matlab. For testing, we have used the two clean databases presented in *Subsection 5.1*. Every image within the database is successively used as query, running the search for this image on the entire database. With each query image, we run seven searches, each with different conditions: no RST, rotation by an angle θ of 30, 60 and 90 degrees, scaling by a factor s of 0.5 and 0.75, and translation by $(\Delta x, \Delta y) = (100,100)$. Similar test is done with the two real condition databases, with no RST.

5.3 Test results

Table 1 summarizes the results of the testing procedure described in *Subsection 5.2*. The recognition accuracy is calculated as the proportion of cases in which the best match corresponds to the query image, with respect to the number of trials. We can see that for clean database and no RST transformations, the system has 100% accuracy.

The average accuracy in case of rotation by an angle θ of 30, 60 and 90 degrees is 96.0 % for pictograms and 97.9 % for pictures. Similarly the average accuracy for scaling, by a factor of 0.5 and 0.75, is 89.8 % for pictograms and 93.1 % for pictures. We can see from the above results that the system performs slightly better for pictures than for pictograms, as pictures have more spectral information than pictograms, which is used in the FMT based RST compensation and in DCT phase matching. The reduction in recognition accuracy compared to no RST transformed images, is from the tolerances of FMT compensation which are passed on to DCT phase matching stage which is sensitive to RST variations.

With real condition database, where the pictograms and pictures are acquired from the camera, the accuracy is 90% for pictograms and 94% for pictures.

6. CONCLUSIONS

A low complexity method for Rotation, Scale and Translation (RST) invariant content-based image retrieval, suitable for a handheld image recognition device was proposed. This method is based on Fourier-Mellin transform (FMT) for RST compensation, Discrete Cosine Transform (DCT) phase match for image recognition and a pre-selection algorithm based on color, for complexity reduction. An efficient method to implement FMT using log-polar grid interpolation is also presented. The image recognition system was tested using 1500 pictograms and 1000 pictures and it showed an average recognition accuracy of 95.2% for pictograms and 96.9% for pictures. Ongoing and future work is in extensive test of the system using real camera acquisition with different real RST conditions.

7. ACKNOWLEDGMENT

REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349-1380, Dec. 2000.

[2] L. Sabeti, Q. M. J. Wu, "New similarity measure for illumination invariant content-based image retrieval", *Proc. of the IEEE International Conference on Automation and Logistics, ICAL'08*, pp. 279-283, Sept. 2008.

[3] D. G. Lowe, "Object recognition from local scale-invariant features", *Proc. of the Seventh IEEE International Conference on Computer Vision, ICCV'99,* vol. 2, pp. 1150-1157, Sept. 1999.

[4] Y. Sheng, H. H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellin descriptors", *Journal of the Optical Society of America A*, vol. 3, no. 6, pp.771 - 776 , 1986.

[5] Chen Qin-Sheng, M. Defrise, F. Deconinck, "Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 16, no. 12, pp. 1156-1168, Dec. 1994.

[6] M. Gueham, A. Bouridane, D. Crookes, O. Nibouche, "Automatic Recognition of Shoeprints using Fourier-Mellin Transform", *Proc. of the third NASA/ESA Conference on Adaptive Hardware and Systems, AHS '08*, pp. 487-491, June 2008.

[7] T. V. Hoang, S. Tabbone, "A Geometric Invariant Shape Descriptor Based on the Radon, Fourier, and Mellin Transforms", *Proc of the 20th International Conference on Pattern Recognition (ICPR'10)*, pp. 2085-2088, 23-26 Aug. 2010.

[8] J. Bracamonte, M. Ansorge, F. Pellandini, P.-A. Farine, "Efficient compressed domain target image search and retrieval", *Proc. of the fourth International Conference on Image and Video Retrieval, CIVR'05*, Singapore, pp. 154-163, July 2005.

[9] P. Ayyalasomayajula, S. Grassi, P.-A. Farine, "Rotation, scale and translation invariant image retrieval method based on circular segmentation and color density", Submitted to ISPA 2011.

[10] B. S. Reddy, B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. on Image Processing,* vol. 5, no. 8, pp. 1266-1271, Aug. 1996.

[11] http://www.mayer-johnson.com

[12] J. Z. Wang, Li Jia, G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 23, no. 9, pp. 947-963, Sep. 2001.

[13] P. Ayyalasomayajula, S. Grassi, P.-A. Farine, "Low complexity image recognition algorithm for handheld applications", *Proc. of the 7th Conference on Ph.D. Research in Microelectronics and Electronics*, PRIME 2011, Italy, July 2011.