Invariant Object Representation with Modified Mellin-Fourier Transform

ROUMEN KOUNTCHEV  VLADIMIR TODOROV  ROUMIANA KOUNTCHEVA
Deptartment of Radio Communications  T&K Engineering  T&K Engineering
Technical University of Sofia  Mladost 3  Mladost 3
BULGARIA  Sofia 1712  Sofia 1712
rkountch@tu-sofia.bg  todorov_vl@yahoo.com  kountcheva_r@yahoo.com

Abstract: - In this paper is presented a method for invariant 2D object representation based on the Mellin-Fourier Transform (MFT), modified for the application. The so obtained image representation is invariant against 2D rotation, scaling, and translation change (RST). The representation is additionally made invariant to significant contrast and illumination changes. The method is aimed at content-based object retrieval in large databases. The experimental results obtained using the software implementation of the method proved the method efficiency. The method is suitable for various applications, such as detection of children sexual abuse in multimedia files, search of handwritten and printed documents, etc.

Keywords: - Modified Mellin-Fourier Transform, 2D object representation, Object retrieval, RST-invariant object representation.

1 Introduction
The development of content-based image retrieval (CBIR) systems requires the creation of new and efficient methods for invariant object representation in still images. The basic methods for invariant object representation with respect to 2D rigid transformations (combinations of rotation, scaling, and translation, RST) are given in significant number of scientific publications [1, 2, 3]. Accordingly, 2D objects in the still grayscale image are depicted by descriptors of two basic kinds: “shape boundary” and “region”. To the first kind (shape boundary) are assigned the chain codes, Fourier descriptors (for example, cumulative angular function and elliptic descriptors); Generalized Hough Transform and Active Shape Model [4]. The skeleton of a shape can be derived by the Medial Axis Transform [5]. To the second kind (“region”) are assigned some geometric characteristics, such as for example: area, perimeter, compactness, dispersion, eccentricity, etc., zero- and first-order statistical moments, centre of gravity, normalized central moments, seven rotation-invariant moments, Zernike polynomial rotation- and scale-invariant, affine transform invariant in respect to position, rotation and different scales along the co-ordinate axes, co-occurrence texture descriptor, etc. The histogram descriptor is proved to be robust to changes of object’s rotation, scale, and partial changes in the viewing direction. The structural information however is lost in the histogram. To solve this problem, the combination of Discrete Wavelet Transform (DWT) or Discrete Fourier Transform (DFT) with the feature extraction method is proposed. For the extraction of the rotation-scale-translation (RST) - invariant features are developed descriptors, based on the log-polar transform (LPT) used to convert rotation and scaling into translation [6] and on the 2D Mellin-Fourier Transform (2D-MFT) [3, 7, 8]. As it is known, the modules of the spectrum coefficients, obtained using the 2D-MFT, are invariant with respect to the RST-transforms of the 2D objects in the image. The basic problem for the creation of the RST-invariant descriptors, in this case is the large number of the calculated spectrum coefficients [11, 12]. With regard to the necessity to reduce their number, and respectively – the time needed for their calculation without decreasing the objects description accuracy, should be solved significant number of problems, regarding the choice of the most informative MFT coefficients and the way of creating the corresponding vector descriptor.

In the paper is presented a method for RST-invariant image representation based on the Modified 2D MFT. In result of the processing
each image is represented by an individual vector, which is then used for content image retrieval in large databases. The paper is arranged as follows: in Section 2 is presented the MFT method, modified for the application; in Section 3 are given some experimental results and Section 4 is the conclusion.

2 Invariant Object Representation with Modified Mellin-Fourier Transform

The method for 2D object representation is aimed at the preparation of a vector description of the object, framed by a square window. The description should be invariant to 2D rotation (R), scaling (S), translation (T) and contrast (C) changes. As a basis for the RSTC description is used the discrete 2D Modified Mellin-Fourier Transform (2D-MMFT). As it is known, the Mellin-Fourier Transform comprises DFT, Log-pol transform (LPT) and DFT again.

The approach, presented below, is aimed at digital halftone images, and comprises the following stages:

1. The pixels \( B(k,l) \) of the original halftone image of size \( M \times N \) are transformed into bi-polar:
   \[
   L(k,l) = B(k,l) - (B_{\text{max}} + 1)/2
   \]
   for \( k = 0,1,\ldots,M-1 \) and \( l = 0,1,\ldots,N-1, \)
   where \( B_{\text{max}} \) is the maximum pixel brightness value.

2. The image is processed with 2D Discrete Fourier Transform (2D-DFT). The Fourier matrix is of size \( n \times n \) (\( n \) - even number). The value of \( n \) defines the size of the window, used to select the object image.

For the invariant object representation are used the complex 2D-DFT coefficients, calculated in accordance with the relation:

\[
F(a,b) = \sum_{k=0}^{n-1} \sum_{l=0}^{n-1} L(k,l) \exp\{-j[2\pi(n/2)(ka + lb)]\}
\]

for \( a = 0,1,\ldots,n-1 \) and \( b = 0,1,\ldots,n-1. \)

The transform comprises two consecutive operations: one-dimensional transform of the pixels \( L(k,l) \), first - for the rows and after that - for the columns of the object image. Since:

\[
\exp\{-j[2\pi(n/2)lb/n]\} = \cos[2\pi(lb/n)] - j\sin[2\pi(lb/n)]
\]

and

\[
\exp\{-j[2\pi(n/2)ka/n]\} = \cos[2\pi(ka/n)] - j\sin[2\pi(ka/n)]
\]

the 2D-DFT is performed as two consecutive one-dimensional DFTs:

- For the fixed values of \( k = 0,1,\ldots,n-1 \) and using the 1D-Fast Fourier Transform (1D-FFT) are calculated the intermediate spectrum coefficients:
  \[
  F(k,b) = \sum_{l=0}^{n-1} L(k,l) \exp\{-j[2\pi(lb/n)]\} = \sum_{l=0}^{n-1} L(k,l) \cos[2\pi(lb/n)] - j \sum_{l=0}^{n-1} L(k,l) \sin[2\pi(lb/n)]
  \]

- For \( b = 0,1,\ldots,n-1 \) and using the 1D-FFT again, are calculated the final Fourier coefficients:
  \[
  F(a,b) = \sum_{k=0}^{n-1} F(k,b) \exp\{j[2\pi(ka/n)]\} = \sum_{k=0}^{n-1} F(k,b) \cos[2\pi(ka/n)] - j \sum_{k=0}^{n-1} F(k,b) \sin[2\pi(ka/n)]
  \]

where \( A_F(a,b) \) and \( B_F(a,b) \) are the real and the imaginary components of \( F(a,b) \) correspondingly.

3. The Fourier coefficients are then centred in accordance with the relation:

\[
F_0(a,b) = F(a - \frac{n}{2}, b - \frac{n}{2})
\]

for \( a, b = 0,1,\ldots,n-1. \)

4. For the next operations some of the Fourier coefficients are retained in accordance with the rule:

\[
F_{\text{retained}}(a,b) = \begin{cases} F_0(a,b), & \text{if } (a,b) \in \text{retained region;} \\ 0, & \text{in all other cases.} \end{cases}
\]

The retained coefficients’ area is a square with a side \( H \leq n \), which envelops the centre \((0,0)\) of the spectrum plane (\( H \) - even number). For \( H < n \) and \( a, b = -(H/2), -(H/2) + 1,\ldots,-1,0,1,\ldots,(H/2) - 1 \) this square contains low-frequency coefficients only.

5. The modules and phases of the coefficients

\[
F_{\text{retained}}(a,b) = D_{\text{retained}}(a,b)e^{j\varphi_{\text{retained}}(a,b)}
\]

are calculated:

\[
D_{\text{retained}}(a,b) = \sqrt{[A_{\text{retained}}(a,b)]^2 + [B_{\text{retained}}(a,b)]^2}
\]

\[
\varphi_{\text{retained}}(a,b) = \arctg[B_{\text{retained}}(a,b)/A_{\text{retained}}(a,b)]
\]

6. The modules \( D_{\text{retained}}(a,b) \) of the Fourier coefficients \( F_{\text{retained}}(a,b) \) are normalized in accordance with the relation:

\[
D(a,b) = p \ln D_{\text{retained}}(a,b)
\]

where \( p \) is the normalization coefficient.

7. The coefficients \( D(a,b) \) are processed with Log-Polar Transform (LPT). The centre \((0,0)\) of the
polar coordinate system \((\rho, \theta)\) coincides with the centre of the image of the Fourier coefficients’ modules \(D(a, b)\) (in the rectangular coordinate system). The transformation of coefficients \(D(a, b)\) from the rectangular \((a, b)\) into the polar \((\rho, \theta)\) coordinate system is performed changing the variables in accordance with the relations:

\[
\rho = \log \sqrt{a^2 + b^2}, \quad \theta = \arctg(b/a)
\]  

(10)

The coordinate change from rectangular into polar is quite clear in the continuous domain, but in the discrete domain the values of \(\rho\) and \(\theta\) should be discrete as well.

Since \(a\) and \(b\) can only have discrete values in the range:

\[
a, b = -(H/2), -1, 0, 1, \ldots, (H/2) - 1,
\]

some of the coefficients \(D(\rho, \theta)\) will be missing. At the end of the transform, the missing coefficients \(D(\rho_i, \theta_i)\) are interpolated using the closest neighbours \(D(a, b)\) in the rectangular coordinate system \((a, b)\) in horizontal or vertical direction (zero-order interpolation).

The number of discrete circles in the polar system with radius \(\rho\) is equal to the number of the discrete angles \(\theta\), for \(i = 1, 2, \ldots, H\). The size (in rectangular coordinates) of the side of the square \(H\) inscribed in the LPT matrix is calculated so, that to ensure maximum part of the coefficients to be transferred without change. For this, the LP transform is modified in accordance with Fig. 1, calculating the radius of the circumscribed circle in correspondence to the relation:

\[
r = \left(\frac{\sqrt{2}}{2}\right)H.
\]  

(11)

The smallest step \(\Delta \rho\) between two concentric circles (the most inside) is calculated:

\[
\Delta \rho = r^{(1/H)}
\]  

(12)

As a result, for the discrete radius \(\rho_i\) and angle \(\theta_i\) for each circle are obtained the relations:

\[
\rho_i = (\Delta \rho)^i = r^{i/H} \quad \text{for } i = 1, 2, \ldots, H,
\]

(13)

\[
\theta_i = (2\pi/H)i \quad \text{for } i = (-H/2), -1, 0, \ldots, (-H/2) - 1.
\]  

(14)

Thus, instead of the logarithmic relation used in the famous LP transform to set the values of the magnitude bins (radiuses) in Step 7 here is used the operation \textit{rising on a power}.

The so modified LP transform we called Exponential Polar Transform, EPT.

After the EPT and the interpolation of the \(D(a, b)\) coefficients is obtained one new, second matrix, which contains the coefficients \(D(x, y)\) for \(x, y = 0, 1, 2, \ldots, H - 1\).

8. The second 2D-DFT is performed for the matrix with coefficients \(D(x, y)\), in accordance with the relation:

\[
S(a, b) = \frac{1}{H^2} \sum_{x=0}^{H-1} \sum_{y=0}^{H-1} D(x, y) \exp \{-j(2\pi/H)(xa + yb)\}
\]  

(15)

for \(a = 0, \ldots, H - 1\) and \(b = 0, \ldots, H - 1\).

The second 2D-DFT is performed in correspondence with Eqs. 3, 4.

9. The modules of the complex coefficients \(S(a, b)\) are then calculated:

\[
D_S(a, b) = \sqrt{|A_S(a, b)|^2 + |B_S(a, b)|^2}
\]  

(16)

where \(A_S(a, b)\) and \(B_S(a, b)\) are correspondingly the real and the imaginary component of \(S(a, b)\).

With this operation the Modified MFT is finished. The processing then continues in the next step with one more operation, aimed at achieving the invariance against contrast changes. In result is obtained the RSTC invariant object representation.

10. The modules \(D_S(a, b)\) of the Fourier coefficients \(S(a, b)\) are normalized:

\[
D_{S_0}(a, b) = \frac{B_{max}[D_S(a, b)]}{D_{S_{max}}(a, b)},
\]  

(17)

where \(D_{S_{max}}(a, b)\) is the maximum coefficient in the matrix \(D_S(a, b)\).

11. The vector for the RSTC-invariant object representation is generated. The vector components are calculated in accordance with the relations:

\[
\Sigma_1(a) = \sum_{b=0}^{H/2} D_{S_0}(a, b), \text{ for } a = 0, 1, \ldots, (H/2) - 1;
\]  

(18)

\[
\Sigma_2(b) = \sum_{a=0}^{H/2} D_{S_0}(a, (H-1-b))
\]  

(19)

for \(b = 0, 1, \ldots, (H/2) - 1\).
From Eqs. 18 and 19 is obtained the RSTC-invariant vector of size $1 \times H$,
\[
\tilde{V} = [\Sigma_1(0), \Sigma_1(1), \ldots, \Sigma_1(\frac{H}{2}-1), \Sigma_2(0), \Sigma_2(1), \ldots, \Sigma_2(\frac{H}{2}-1)]^T
\]  
which comprises two vectors of size $1 \times H/2$:
\[
\tilde{V}_a = [\Sigma_1(0), \Sigma_1(1), \ldots, \Sigma_1(\frac{H}{2}-1)]^T
\]
\[
\tilde{V}_b = [\Sigma_2(0), \Sigma_2(1), \ldots, \Sigma_2(\frac{H}{2}-1)]^T .
\]

### 3 Search of Closest Objects in Image Databases

The search of closest objects in image databases (DB) for the image request is based on the detection of the minimum Euclidean distance $d_E$ between their RSTC vectors. For two $H$-dimensional vectors $\tilde{V}_i$ and $\tilde{V}_j$ this distance is defined by the relation:
\[
d_E(\tilde{V}_i, \tilde{V}_j) = \sum_{m=0}^{H-1} [v_i(m) - v_j(m)]^2 ,
\]
where $v_i(m), v_j(m)$ are the $m^{th}$ components of vectors $\tilde{V}_i, \tilde{V}_j$ for $i \neq j$. The decision for the image request classification, represented by the $H$-dimensional RSTC-vector $\tilde{V}$ is taken on the basis of the preset image classes in the DB and their RSTC vectors $\tilde{V}_{\alpha \in C_\beta}$ for $\alpha = 1, 2, \ldots, P$ and $\beta = 1, 2, \ldots, Q$ (P – the number of vectors in the DB, which describe the objects in the class $C_\beta$, and Q is the number of classes). Then, a classification rule is applied, based on the $K$-nearest neighbours and “majority vote” algorithms [9, 10]:
\[
\tilde{V} \in C_\beta, \text{ if } d_1(\tilde{V}, \tilde{V}_{\alpha_1}) \leq d_2(\tilde{V}, \tilde{V}_{\alpha_2}) \leq \ldots \leq d_K(\tilde{V}, \tilde{V}_{\alpha_K})
\]
where: $K$ is an odd number; $\alpha_k$ and $\beta_k$ for $k=1, 2, \ldots, K$ and $\beta$ are correspondingly in the intervals: $[1, P]$ - for $\alpha_k$, and $[1, Q]$ - for $\beta_k$ and $\beta$. The class $C_\beta$ of the vector $\tilde{V}$ is defined by the most frequent value $\beta$ of the indices $\beta_k$ of the vectors $\tilde{V}_{\alpha_k}$.

### 4 Experimental Results

For the experiments was used the software implementation of the method in C++, Windows environment.

Significant part of the experiments aimed to prove the efficiency of the Modified Mellin-Fourier Transform. For the experiments was used the well-known test image “Lena”, $256 \times 256$ pixels, 8 bpp. The experiments were performed for various values of the main parameters: the side of the subscribed circle, the number of discrete radiuses, etc.

On Fig. 2.a is shown the experimental image “Lena”, on which are indicated the points, which participate in the EPT (the black points are not retained and the image is restored after corresponding interpolation). On Fig. 2.b are shown the points, which participate in the LPT (the used points here are marked as black). This experiment confirms the efficiency of the new approach, because in the well-known LPT the retained central part of the processed image is smaller.

Some of the experimental results performed with the same test image “Lena”, are given below.

On Fig. 3.a,b are shown the original test image and the result obtained after DFT, when the size of the window of selected coefficients (H) was 96.

On Fig. 4.a, b are shown the original test image and the result obtained after DFT, when the size of the window of selected coefficients (H) was 96.
Fig. 4. The first 11 closest images from a database of 180 faces.

The results obtained confirm the RST-invariance of the method representation: the first 5 images are of the test image “Lena” – the first two are scaled up and cropped; the next three are the original, and the same image rotated on 180° and 90° correspondingly.

The experiments aimed at detection of children sexual abuse in multimedia files need additional preprocessing. For this, color image segmentation was first performed, in order to detect naked parts of human bodies and then these parts were extracted from the images and defined as individual objects. After that the object search in the corresponding database was initiated. Special attention was paid to ability for children and adults faces recognition. The experiments confirm that this recognition is successful enough. On Fig. 6 below are shown some of the results obtained for search of child’s face in a mixed database, containing 200 faces of children and adults. Each image in the database was classified as belonging to one of these two classes.

Fig. 6. Results obtained for image request (upper left) in a mixed database of 200 faces (children and adults)

The images on Fig. 6 are the closest 11 to the image request (upper-left) in the test database. On the same row are arranged the closest 5; the next 6 are arranged on the second row, from left to right. The experiment confirmed the method reliability when the searched face is of same person: the second image in the first row is of the same child as the image request. The situation is same with images 5, and 7 in spite of the fact that in the database were included photos of more than 40 children. The error in this search result is one face only – the third from the left in the lower row. In some cases is possible to get large number of wrong images in the selection. In order to solve possible uncertainties, the final decision is taken in correspondence with Eq. 24. The detailed presentation of the decision rules used for the analysis is not object of this work.

The database of scanned documents, comprised images of scanned texts, and signatures (Q=2). The database contained more than 100 samples (P>100) of each class; texts comprised examples of Latin and Cyrillic alphabets, printed and handwritten texts. All images were of size 256×256 pixels, greyscale (8 bpp, n=256); Bmax=255 and p=16. The experiments were performed for two versions of vector generation: 1st version: the size of the retained coefficients square (H) and the vector size - equal to 96; 2nd version: the size of the retained coefficients square (H) and the vector size - equal to 128. In Figs. 7 and 8 are shown results obtained for one of the test images (the image request is the upper left one) and the closest K=11 images. In most cases (90%) the information provided by the short vector was enough for the right classification, but for some test images we had small number of mistakes. In Fig. 7 is given the result for the short vector (version 1). There are 3 mistakes, i.e. the last 3 images (signatures) were classified as text, instead as belonging to the class of signatures. As it is seen in Fig. 8, the use of the longer vector ensured the right decision (there are no mistakes at all). One of the
images in the DB (the second in Figs. 7, 8) was the same as the image request, but with changed contrast. In both experiments it was qualified as closest, which proves the method invariance to contrast changes.

![Fig. 7. Results for vectors of 96 coefficients](image1)

![Fig.8. Results for vectors of 128 coefficients](image2)

6 Conclusion
In this paper is presented a method for invariant object representation with Modified MFT. The main differences from the famous MFT are: a) the first DFT is performed for limited number of coefficients only. In result is obtained an approximated image representation, suitable for the object representation; b) instead of the Log-Pol Transform, here was used the Exponential-Polar Transform (EPT), in accordance with the description in Step 7 of the algorithm. As a result, the part of the participating points from the matrix of the Fourier coefficients’ modules is larger (i.e. bigger central part of the image participates in the EPT and correspondingly – in the object description). The number of coefficients, used for the object representation is additionally limited in accordance with the vector length selection (vector representation comprises two vectors of size $1 \times H/2$ only). Besides, the new transform is invariant to contrast changes because of the normalization performed in Step 10. In result, the MMFT described above has the following advantages over the MFT:

- The number of transform coefficients used for the object representation is significantly reduced and from this naturally follows the lower computational complexity of the method, which permits real-time applications.
- The choice of coefficients, used for the vector calculation offers wide possibilities by setting large number of parameters, each of relatively wide range, which permits the method use in various applications: content- and context-based object retrieval in image databases, face recognition, etc.

The new approach, presented in this work, permits reliable object detection and identification in various positions, lighting conditions and view points.

Acknowledgements
This work was supported by the National Fund for Scientific Research of the Bulgarian Ministry of Education and Science, Contract VU-I 305.

References