RSTC-Invariant Object Representation with 2D Modified Mellin-Fourier Transform

ROUMEN KOUNTCHEV
Deptartment of Radio Communications
Technical University of Sofia
BULGARIA
rkountch@tu-sofia.bg
www.tu-sofia.bg

VLADIMIR TODOROV
T&K Engineering
Mladost 3
Pob 12
Sofia 1712
BULGARIA
todorov_vl@yahoo.com

ROUMIANA KOUNTCHEVA
T&K Engineering
Mladost 3
Pob 12
Sofia 1712
BULGARIA
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 image databases. A new algorithm for fast closest vector search in the database is proposed as well. 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, Fast closest vector search in a database.

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.

Significant research efforts had recently been aimed at solving the problem with the efficient image retrieval in large databases. In [6] is presented an on-line content-based image retrieval system using joint querying and relevance feedback scheme based on both high-level and low-level features. The color feature extraction is based on specially developed algorithms for color correlogram and autocorrelogram. The so obtained information is used for extracting and indexing low-level features of images. In [7] are proposed two methods for retrieving texture images. In the first method, discrete wavelet transform and gradient operation are combined to extract features of an image with principal component analysis used to determine weights of individual extracted features, while in the second method, only gradient operation without involvement of discrete wavelet transform was used to extract features. More methods are presented in [8] and [9], based on Fuzzy C-Means and on cross correlation in the frequency domain between input data and the input weights of fast neural networks.

A method for Association-Based Image Retrieval is given in [10], called association-based image retrieval, which tries to mimic human memory.

All these methods and algorithms have their advantages and disadvantages. It is not possible to develop an algorithm, which to be equally efficient for all kinds of images and objects representation or search. Besides, additional difficulties appear in the case, when the searched objects could be represented in multiple views. 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 [11] and on the 2D Mellin-Fourier Transform (2D-MFT) [3,12,13,16]. 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 [14,15]. 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 this 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:

Step 1. The pixels B(k,l) of the original halftone image of size M×N are transformed into bi-polar:

$$L(k,l) = B(k,l) - (B_{max}+1)/2 \qquad (1)$$

for k = 0,1,...,M-1 and l = 0,1,.., N-1,

where B_{max} is the maximum value in the pixel quantization scale.

Step 2. The image is processed with 2D Discrete Fourier Transform (2D-DFT). The Fourier matrix is of size n×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:

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

for $a = 0,1,..,n-1$ and $b = 0,1,..,n-1$.

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:

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

and

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

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

- For the fixed values of $k = 0,1,..,n-1$ and using the 1D-Fast Fourier Transform (1D-FFT) are calculated the intermediate spectrum coefficients:

$$F(k,b) = \sum_{l=0}^{n-1} L(k,l)\exp\{-j[2\pi(lb/n)]\} = \qquad (3)$$

$$= \sum_{l=0}^{n-1} L(k,l)\cos[2\pi(lb/n)] - j\sum_{l=0}^{N-1} L(k,l)\sin[2\pi(lb/n)]$$

- For $b = 0,1,...,n-1$ and using the 1D-FFT again, are calculated the final Fourier coefficients:

$$F(a,b) = \sum_{k=0}^{n-1} F(k,b)\exp\{- j[2\pi(ka/n)]\} = \\ = \sum_{k=0}^{n-1} F(k,b)\cos[2\pi(ka/n)] - \\ - j\sum_{k=0}^{n-1} F(k,b)\sin[2\pi(ka/n)] = \\ = A_F(a,b) - jB_F(a,b) \quad (4)$$

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

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

$$F_0(a,b) = F(a-\tfrac{n}{2}, b-\tfrac{n}{2}) \quad (5)$$

for a, b = 0,1,.., n-1.

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

$$F_{0R}(a,b) = \begin{cases} F_0(a,b), \text{if } (a,b) \in \text{retained region;} \\ 0 \quad - \quad \text{in all other cases.} \end{cases} \quad (6)$$

The retained coefficients' area is a square with a side $H \le n$, which envelops the centre (0,0) of the spectrum plane (H - even number). For $H < n$ and

$$a,b = -(H/2), -(H/2)+1,...,-1,0,1,...,(H/2)-1$$

this square contains low-frequency coefficients only.

Step 5. The modules and phases of the coefficients

$$F_{0R}(a,b) = D_{F_{0R}}(a,b)e^{j\varphi_{F0R}(a,b)}$$

are calculated:

$$D_{F_{0R}}(a,b) = \sqrt{[A_{F_{0R}}(a,b)]^2 + [B_{F_{0R}}(a,b)]^2} \quad (7)$$

$$\varphi_{F_{0R}}(a,b) = \text{arctg}[B_{F_{0R}}(a,b)/A_{F_{0R}}(a,b)] \quad (8)$$

Step 6. The modules $D_{F_{0R}}(a,b)$ of the Fourier coefficients $F_{0R}(a,b)$ are normalized in accordance with the relation:

$$D(a,b) = p \ln D_{F_{0R}}(a,b) \quad (9)$$

where p is the normalization coefficient.

Step 7. The coefficients $D(a,b)$ are processed with Log-Polar Transform (LPT). The centre (0,0) of the polar coordinate system (ρ,θ) 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 (ρ,θ) coordinate system is performed changing the variables in accordance with the relations:

$$\rho = \log\sqrt{a^2+b^2}, \ \theta = \text{arctg}(b/a) \quad (10)$$

The coordinate change from rectangular into polar is quite clear in the continuous domain, but in the discrete domain the values of ρ and θ should be discrete as well.

Since **a** and **b** can only have discrete values in the range:

$$a,b = -(H/2),...,-1,0,1,...,(H/2)-1,$$

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 ρ^i is equal to the number of the discrete angles θ_i for i = 1,2,..,H. The number of discrete circles in the polar system with radius ρ^i is equal to the number of the discrete angles θ_i for i = 1,2,..,H. The radius of the circumscribed circle is calculated in correspondence to the relation:

$$r = (\sqrt{2}/2)H. \quad (11)$$

The smallest step Δρ between two concentric circles (the most inside) is calculated:

$$\Delta\rho = r^{(1/H)}. \quad (12)$$

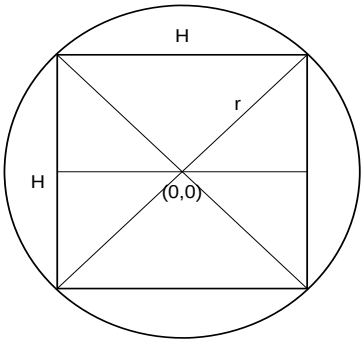


Fig. 1. Geometric relations between r and H

As a result, for the discrete radius ρ_i and angle θ_i for each circle are obtained the relations:

$$\rho_i = (\Delta\rho)^i = r^{i/H} \quad \text{for } i = 1,2,..,H, \quad (13)$$

$$\theta_i = (2\pi / H)i \text{ for } i = (-H/2),.,0,.,(-H/2)-1. \quad (14)$$

Thus, instead of the logarithmic relation used in the famous LP transform to set the values of the magnitude bins (radiuses) in the LPT, here is used the operation *rising on a power*. The so modified LP transform we called Exponential Polar Transform, EPT.

An example of the difference between the LPT and the EPT is shown on Fig. 2 below. 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 much smaller. Besides, this approach permits the user to choose the size of the square and the radius, and in result – to set the retained area size and the density of the used original points.

a. Points retained by EPT; b. Points retained by LPT
Fig.2. Experimental images which indicate the points, retained after EPT and LPT

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,..,H-1.

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

$$S(a,b) = \frac{1}{H^2}\sum_{x=0}^{H-1}\sum_{y=0}^{H-1} D_I(x,y)\exp\{-j[(2\pi/H)(xa+yb)]\} \quad (15)$$

for $a = 0,..,H-1$ and $b = 0,..,H-1$.

The second 2D-DFT is performed in correspondence with Eqs. 3, 4, applying consecutively the 1D-DFT on the rows of the matrix [D] first, and then - on the columns of the intermediate matrix obtained.

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

$$D_S(a,b) = \sqrt{[A_S(a,b)]^2 + [B_S(a,b)]^2} \quad (16)$$

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.

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

$$D_{S_0}(a,b) = B_{max}[D_S(a,b)/D_{Smax}(a,b)], \quad (17)$$

where $D_{Smax}(a,b)$ is the maximum coefficient in the matrix $D_S(a,b)$.

Step 11. The vector for the RSTC-invariant object representation is generated. Two methods for vector representation are proposed in this work.

The First method is based on the calculation of the sums of the spectrum coefficients' $D_{S_0}(a,b)$ values in the rows and columns of the normalized amplitude DFT spectrum (Eq. 17). As a result of the complex-conjugated symmetry of the coefficients in quadrants I - III and II - IV, the summing is performed for the frequencies:

a = 0,1,..,(H/2)-1 and b = 0,1...,(H/2)-1 only.

The vector components for each set of frequencies *a* and *b* are calculated in accordance with the relations:

$$\Sigma_1(a) = \sum_{b=0}^{H/2} D_{S_0}(a,b); \quad (18)$$

$$\Sigma_2(b) = \sum_{a=0}^{H/2} D_{S_0}(a,(H-1-b)) \quad (19)$$

From Eqs. 18 and 19 is obtained the RSTC-invariant vector of size 1×H,

$$\vec{V} = [\Sigma_1(0),\Sigma_1(1),.... \\ ...., \Sigma_1(\frac{H}{2}-1),\Sigma_2(0),\Sigma_2(1),..,\Sigma_2(\frac{H}{2}-1)]^T \quad (20)$$

which comprises two vectors of size 1×H/2:

$$\vec{V}_a = [\Sigma_1(0),\Sigma_1(1),..,\Sigma_1(\frac{H}{2}-1)]^T \quad (21)$$

$$\vec{V}_b = [\Sigma_2(0), \Sigma_2(1), .., \Sigma_2(\frac{H}{2} - 1)]^T \quad . \qquad (22)$$

The Second method is based on the use of coefficients $D_{S_0}(a,b)$ of highest energy in the amplitude spectrum 2D-MFT of size H×H. One example test image (Caltech database "FACES") and its corresponding 2D-MFT spectrum obtained with the software implementation of the method, are shown correspondingly on Fig 3 a, b. For the extraction of the retained coefficients $D_{S_{0R}}(a,b)$ is used the mask, shown on Fig. 3c (the part, coloured in yellow, corresponds to the area of the complex-conjugated coefficients, $D^*_{S_{0R}}(a,b)$). The shape of the mask approximates the area, where the energy of the mean 2D-MFT spectrum is concentrated.

Fig. 3.a. Test image

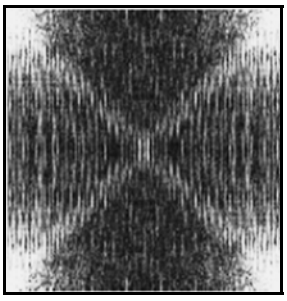


Fig.3.b. The amplitude MFT spectrum of the test image

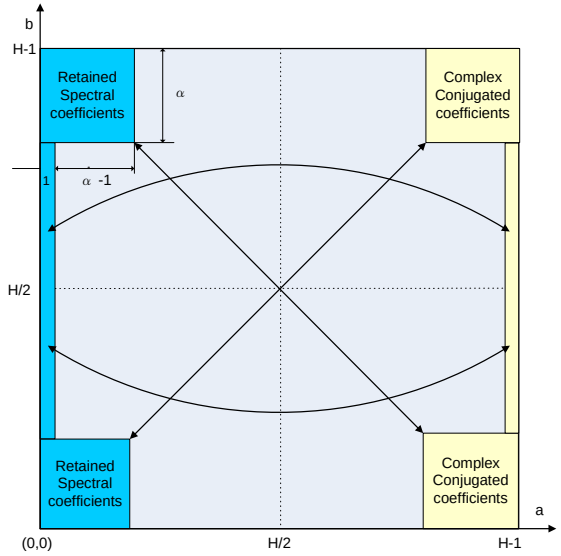


Fig. 3.c. The mask of the retained coefficients of the amplitude 2D-MFT spectrum

The parameter α of the mask defines the number of retained coefficients in correspondence with the relation:

$$R = 2\alpha^2 + H - 2\alpha = 2\alpha(\alpha - 1) + H \, . \qquad (23)$$

The v_m components for m=1,2,..,R of the corresponding RSTC-invariant vector:

$\vec{V} = [v_1, v_2, .., v_R]^T$ are defined by coefficients $D_{S_{0R}}(a,b)$, arranged as one-dimensional massif after lexicographic tracking of the 2D-MFT spectrum in the mask area, colored in blue.

The vector $\vec{V}$, calculated in accordance with the second method is longer than the vector, calculated following the first method, i.e. R > H. In result, the computational complexity in the second case is higher, but the vector contains more information about the represented object and this ensures higher accuracy for the efficient object search.

The block diagram, representing the calculation of the vector for invariant object representation in a square of size n×n in accordance with the method, described above, is given in Fig. 4.

3 Search of Closest Objects in Image Databases

The search of closest objects to the image request in image databases (DB) is based on the detection of the minimum squared Euclidean distance (EUD) $d_E(.)$ between their RSTC-invarant vectors. For two H(R)-dimensional vectors $\vec{V}$ and $\vec{V}_j$ this distance is defined by the relation:

$$d_E(\vec{V}, \vec{V}_j) = \sum_{m=0}^{H(R)-1} [v(m) - v_j(m)]^2, \qquad (24)$$

where $v_i(m), v_j(m)$ are the m^{th} components of vectors $\vec{V}$ and $\vec{V}_j$.

The decision for the image request classification, represented by the H(R)-dimensional RSTC-vector $\vec{V}$ is taken on the basis of the image classes in the DB and their RSTC-invariant vectors

$$\vec{V}^{\beta}_{\alpha_\beta} \in C_\beta \quad \text{for} \quad \alpha_\beta = 1,2,\ldots,P_\beta \text{ and } \beta = 1,2,\ldots,Q$$

P_β - the number of vectors in the DB, which

After that, a classification rule is applied, based on the K-nearest neighbours (K-NN) and "majority vote" algorithms [17, 18, 19, 20]:

$$\vec{V} \in C_{\beta_0}, \text{ if } d_1(\vec{V}, \vec{V}^{\beta_1}_{\alpha_{\beta_1}}) \le d_2(\vec{V}, \vec{V}^{\beta_2}_{\alpha_{\beta_2}}) \le .... \\ ..... \le d_K(\vec{V}, \vec{V}^{\beta_K}_{\alpha_{\beta_K}}) \qquad (25)$$

where K is an odd number;

$d_k(\vec{V},\vec{V}_{\alpha_{\beta_k}}^{\beta_k})$ is the squared Euclidian distance between vectors $\vec{V}$ and $\vec{V}_{\alpha_{\beta_k}}^{\beta_k}$, for k = 1, 2, … , K;

$\vec{V}_{\alpha_{\beta_k}}^{\beta_k}$ is the k^{th} vector with index α_{β_k} from the class C_{β_k}, which is at minimum distance from the query $\vec{V}$. The indices α_{β_k}, β_k and β_0 are correspondingly in the ranges: $[1, P_{\beta_k}]$ - for α_{β_k}, and [1, Q] - for β_k and β_0.

The class C_{β_0} of the vector $\vec{V}$ in Eqs. 25 is defined by the most frequent value β_0 of the indices β_k of the vectors $\vec{V}_{\alpha_{\beta_k}}^{\beta_k}$:

$$\beta_0 = \max\{h(\beta_k)\} \quad \text{for } \beta_k = 1,2,..,Q \text{ and } k=1,2,..,K. \tag{26}$$

Here $h(\beta_k)$ is the histogram of the indices β_k, for the relations in Eqs. (25).

4. Fast Search of Closest Vector in the Image DB

In order to perform the search of closest vector to the vector request $\vec{V}$ in the DB comprising $J = Q\sum_{\beta=1}^{Q} P_\beta$ vectors, could be used the approach, described in [19], and based on the representation of the squared Euclidean distance EUD between any two n-dimensional vectors, $\vec{V}$ and $\vec{V}_j$ (j=1,2,..,J) by the relation:

$$\begin{aligned} d_E(\vec{V},\vec{V}_j) &= \sum_{m=0}^{n-1}[v(m)-v_j(m)]^2 = \\ &= \left\|\vec{V}\right\|^2 + \left\|\vec{V}_j\right\|^2 - 2\sum_{m=0}^{n-1} v(m)\times v_j(m), \end{aligned} \tag{27}$$

where the size of the vectors is correspondingly

n = H(R),

and

$$\left\|\vec{V}\right\|^2 = \sum_{m=0}^{n-1}[v(m)]^2,\ \left\|\vec{V}_j\right\|^2 = \sum_{m=0}^{n-1}[v_j(m)]^2.$$

The modules of the vectors $\vec{V}$ and $\vec{V}_j$ in the equation above are not related to $d_E(\vec{V},\vec{V}_j)$, but their scalar product is related as follows:

$$f(j) = \sum_{m=0}^{n-1} v(m)\times v_j(m).$$

The vectors $\vec{V}_j$ (j=1,2,..,J) are calculated in advance for the images used as training set for each image class and are stored as additional information (metadata) in the DB. In case, that the function f(j) has a maximum for some value of the variable $j = j_0$, the corresponding distance $d_E(\vec{V},\vec{V}_{j0})$ is minimum. For the exact detection of j_0 is accepted the value, for which j_0 = min. The closest vector $\vec{V}_{j_0}$ in the DB to the vector request $\vec{V}$ could be defined in accordance with the rule:

$$\begin{aligned} &d_E(\vec{V},\vec{V}_{j0}) = \min, \\ &\text{if } \left|\sum_{m=0}^{n-1} v(m)\times v_{j_0}(m)\right| = \max \\ &\text{for } j_0 = 1,2,..,J. \end{aligned} \tag{28}$$

Additional acceleration for the calculations in Eq. 28 is achieved, when the following suggestions are taken into account:

- In case, that $v(m) \ge 0$ and $v_j(m) \ge 0$ the following relation is satisfied:

$$\sum_{m=0}^{n-1} v(m)\times v_j(m) < v_{max}\sum_{m=0}^{n-1} v_j(m), \tag{29}$$

where

$$v_{max} = \max\{v(m)\} \text{ for } m=0,1,..,n-1.$$

Then, for the so-called "modified" squared Euclidean distance $D_E(\vec{V}_i,\vec{V}_j)$ (MEUD) follows:

$$\begin{aligned} &D_E(\vec{V},\vec{V}_j) = \\ &= \left\|\vec{V}\right\|^2 + \left\|\vec{V}_j\right\|^2 - 2v_{max}\sum_{m=0}^{n-1} v_j(m) \le d_E(\vec{V},\vec{V}_j). \end{aligned} \tag{30}$$

Then, Eq. 28 could be transformed as follows:

$$\begin{aligned} &d_E(\vec{V},\vec{V}_{j0}) = \min, \\ &\text{if } [v_{max}\sum_{m=0}^{n-1} v_{j_0}(m)] = \max \text{ for } j_0 = 1,2,..,J. \end{aligned} \tag{31}$$

• In case, that the components v(m) and $v_j(m)$ have positive and negative values, they should be transformed in such a way, that to have positive values only:

$$v'(m) = v(m) + \min\{v(m), v_j(m)\}, \quad (32)$$

$$v'_j = v_j + \min\{v(m), v_j(m)\} \quad (33)$$

$$\text{for } m = 1,2,..,n-1.$$

In this case, for the modified distance $D_E(\vec{V},\vec{V}_j)$ is obtained:

$$D_E(\vec{V},\vec{V}_j) = \\ = \left\|\vec{V}\right\|^2 + \left\|\vec{V}_j\right\|^2 - 2v'_{max}\sum_{m=0}^{n-1} v'_j(m) \le d_E(\vec{V},\vec{V}_j), \quad (34)$$

and the rule, represented by Eq. 31 is changed accordingly:

$$d_E(\vec{V},\vec{V}_{j_0}) = \min, \\ \text{if } [v'_{max}\sum_{m=0}^{n-1} v'_{j_0}(m)] = \max \text{ for } j_0 = 1,2,..,J. \quad (35)$$

where $v'_{max} = \max\{v'(m)\}$ for $m = 0,1,..,n-1$.

The rule for the classification of the vector $\vec{V}$ (Eq. 25), based on the K-NN and MEUD is transformed as follows:

$$\vec{V} \in C_{\beta_0}, \text{ if} \quad (36)$$

$$S_1(\vec{V},\vec{V}_{\alpha_{\beta_1}}^{\beta_1}) \ge S_2(\vec{V},\vec{V}_{\alpha_{\beta_2}}^{\beta_2}) \ge .... \ge S_K(\vec{V},\vec{V}_{\alpha_{\beta_K}}^{\beta_K}),$$

where

$$\beta_0 = \max\{h(\beta_k)\}$$

$$\text{for } \beta_k = 1,2,.., Q \text{ and } k = 1,2,..,K;$$

$$S_k(\vec{V},\vec{V}_{\alpha_{\beta_k}}^{\beta_k}) = [v'_{max}\sum_{m=0}^{n-1} v'^{\beta_k}_{\alpha_{\beta_k}}(m)] \quad (37)$$

Comparing the Eqs. 25, 27 and 28 with Eqs. 36, 31 and 35 respectively follows, that the use of the similarity criterion $S_k(\vec{V},\vec{V}_{\alpha_{\beta_k}}^{\beta_k})$ instead of $d_k(\vec{V},\vec{V}_{\alpha_{\beta_k}}^{\beta_k})$ does not influence the classification results for the vector $\vec{V}$, but the number of multiplications is reduced (n.J) times, with retained number of sums. Besides, the corresponding sum $\sum_{m=0}^{n-1} v'^{\beta_k}_{\alpha_{\beta_k}}(m)$ for each vector $\vec{V}_{\alpha_{\beta_k}}^{\beta_k}$ could be defined in advance and saved in the DB metadata. In result, the computational complexity is significantly reduced and the classification of the vector-request $\vec{V}$ - enhanced.

In order to improve the similarity arrangement for the first K-nearest vectors to the vector-request, in Eq. 36 is possible to use the criterion "Cosine similarity" (CSim) [15], instead of $S_k(.)$. In correspondence with this criterion, the nearness between the two vectors, $\vec{V}$ and $\vec{V}_j$ could be evaluated using the relation:

$$CSim(\vec{V},\vec{V}_j) = \quad (38)$$

$$= \frac{\sum_{m=0}^{n-1} v'(m) \times v'_j(m)}{\left[\sqrt{\sum_{m=0}^{n-1}[v'(m)]^2}\right] \times \left[\sqrt{\sum_{m=0}^{n-1}[v'_j(m)]^2}\right]}$$

In order to avoid "square root" operations, Eq. 38 could be transformed as Squared Cosine Similarity (SCSim) = $(CSim)^2$. Further reduction of the needed calculations is achieved using the Modified Squared Cosine Similarity (MSCSim), defined as follows:

$$MSCSim(\vec{V},\vec{V}_j) = \quad (39)$$

$$= \frac{\left[v'_{max} \times \sum_{m=0}^{n-1} v'_j(m)\right]^2}{\left[\sum_{m=0}^{n-1}[v'(m)]^2\right] \times \left[\sum_{m=0}^{n-1}[v'_j(m)]^2\right]}$$

Then, in Eq. 36 the term $S_k(.)$ is substituted by MSCSim(.). In result, the number of multiplications, needed for the calculation of MSCSim is reduced n times, and the search of the closest vector in a DB, containing J vectors, to the vector request, is enhanced (n.J) times. The efficiency of the presented approach for enhanced search of closest vector grows up with the increasing of the number of vectors in the DB.

4 Experimental Results

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

Significant part of the experiments was aimed to prove the efficiency of the Modified Mellin-Fourier Transform. For the experiments was used the well-known test image "Lena", 256 × 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.

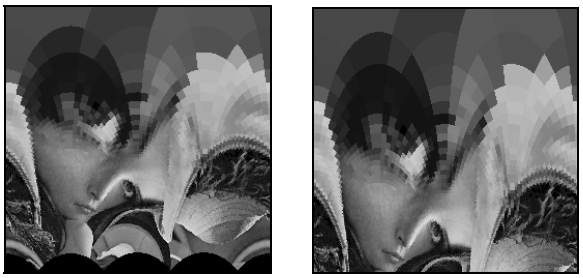
Some of the experimental results performed with the test image "Lena", are given below.

On Fig. 5.a,b are shown the original test image and the result obtained after DFT, when the size of the window of retained coefficients (H) was 96.

a b.

Fig. 5.a) original test image "Lena"; b) result, after DFT for H=96.

On Fig. 6.a,b are shown the results obtained after EPT for two values of the window of selected coefficients (H) size: 96 and 64 correspondingly.



a b

Fig. 6. a) The test image "Lena" after EPT with parameters H=96; r=96; b) The test image "Lena" after EPT with parameters H=64; r=64.

The next part of the experiments was aimed at the content-based object retrieval. For this were used 3 specially developed image databases of the Technical University - Sofia: the first contained 180 faces of adult people, the second – 200 faces of adult people and children and the third – more than 200 scanned documents. Most of the faces in the databases are cropped from larger images. These photos were taken in various lighting conditions with many shadows, different views, etc.

Very good results were obtained for search of similar faces in the databases. In the test database of adults, was included the image "Lena" rotated in 90^0, and 270^0 and scaled up cropped part of the same original test image. The experiments proved the method efficiency. The image request (one of the cropped images from the test image "Lena") was the upper-left one in Fig. 7 below. In Fig. 7 are also shown the first 11 closest images from the database. All images are of size 256×256 pixels, greyscale (8 bpp, n=256). The first 5 closest are arranged in the first row, from left to right; the remaining 6 are arranged in the second row, from left to right. Naturally, the distance between vector request and vectors, corresponding to faces, different from "Lena" is larger than these for the first 4 images. The experiments were performed under following conditions: B_{max}=255; K=16; the size of the retained coefficients square (H) is 128. The vector is calculated in accordance with the First method (Eq. 20). The vector size in this case is 128.

Fig. 7. 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 180^0 and 90^0 correspondingly. The distances between vectors, representing the image request and the first closest ones in the DB (belonging to same person) have minimum values, but the next ones naturally are placed at larger distances.

The experiments aimed at detection of children sexual abuse in multimedia files need additional pre-processing. 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. 8 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.

The experiments were performed under following conditions: B_{max}=255; K=16; the size of the retained coefficients square (H) is 128. The vector is calculated in accordance with the Second method (Eq. 23). In this case α=20 and the vector size is R=888.

The images on Fig. 8 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.

Fig. 8. Results obtained for image request (upper left) in a database of 200 faces (children and adults).

Special attention was paid for the detection of 3D objects, represented by several multi-view 2D images. For this, was used special database of the Technical University-Sofia, containing more than 200 multi-view 2D images of various 3D objects. All images are of size 256×256 pixels; 24 bpp. Each object was represented by 4 views, placed at equal distances in a sector of 20^0. Another set of 4 images was taken from positions, placed at 5^0 up.

For illustration, here is used a small database of 4 objects (Fig. 9), represented by 30 images. Each object had 4 views. More views (8) were included in the database for the object "Cup" only. Each image represented the object in different scale.

a b

c d

Fig. 9.a) Glass; b) Mug; c) Cup; d) Egg

The experiments proved that the 3D object recognition is reliable in a view angle of 20^0.

On Fig. 10 are shown the results obtained for one of the test images. The image request ("Mug") is at the upper left corner of Fig. 10, and after it on the same row follow the closest images detected in the database (the names of the images are in accordance with their number in the database). The remaining closest images are arranged in the next two rows.

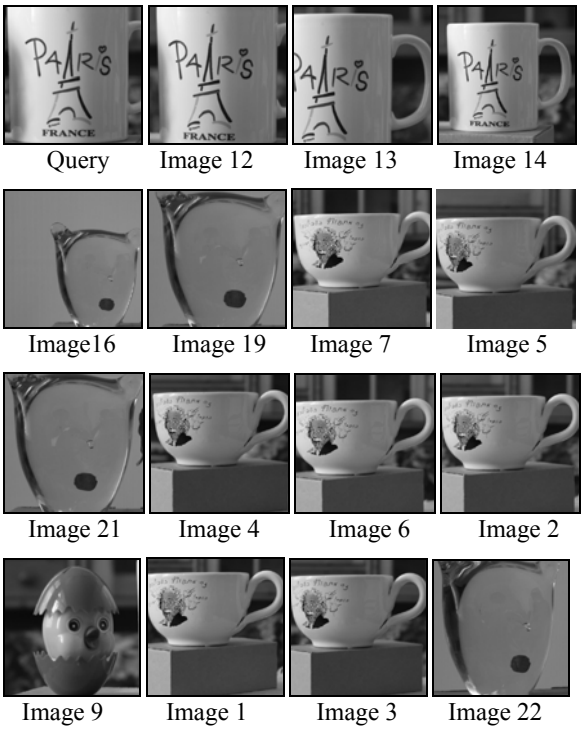


Fig. 10. Results for the image request (upper left) in a database of multi-view images.

The graphic representation of the distances, calculated for the corresponding vectors of the objects represented by the 2D images in the database, is shown on Fig. 11.

It is easy to notice, that the first closest images (No. 1-3, corresponding to images 12, 13 and 14 from Fig. 10) are of the same object. Besides, they are the only images of this object in the test database. The vectors of next closest images are at much

larger distance, which proves their belonging to another object. This example was used for illustration purposes only. In real application tasks, special decision rules should be set, developed in accordance with the objects features.

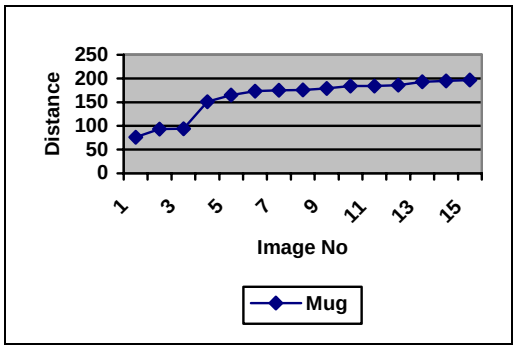


Fig. 11. Distances between the vector of the image request and the vectors of the closest objects in the database.

Another group of experiments was aimed at the analysis of scanned documents. The database of scanned documents, comprised images of scanned texts, and signatures (Q=2). The database contained more than 100 samples (J>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); B_{max}=255 and K=16. The experiments were performed for two versions of vector generation in accordance with the first method (Eq. 20): *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. 12 and 13 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. 12 is given the result for the short vector (H=96). 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. 13, the use of the longer vector (H=128) ensured the right decision (no mistakes at all).

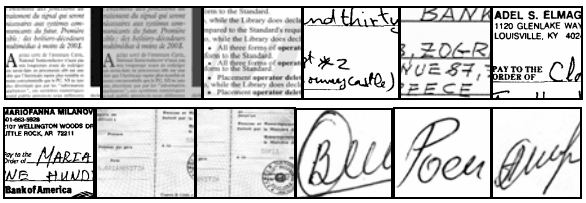


Fig. 11. Results for vectors of 96 coefficients

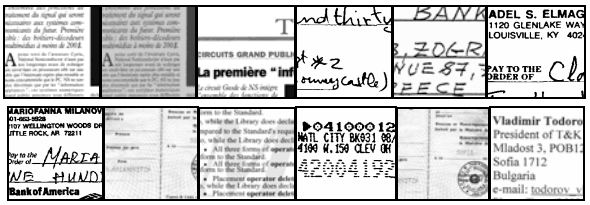


Fig.12. Results for vectors of 128 coefficients

One of the images in the DB (the second in Figs. 12 and 13) 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.

6 Conclusions

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 corresponding 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×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.
- A new, simplified algorithm for fast search of closest vector in large image databases is developed.

❖ The method permits to detect similarity of 3D objects, represented by multiple 2D views, in a view angle of 20^0. Additional experiments show that the view angle could be wider for 3D objects of high symmetry.

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, Youth and Science, Contract VU-I 305.

References

[1] M. Nixon and A. Aguado, *Feature Extraction and Image Processing*, Newness, Oxford, (2002)

[2] L. Costa, R. Cesar, *Shape Analysis and Classification: Theory and Practice*, CRC Press LLC, 2001.

[3] G. Ritter, J. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, 2nd ed., CRC Press LLC, 2001.

[4] T. Cootes, C. Taylor, D. Cooper, and J. Graham, *Active Shape Models - their Training and Application*, CVIU, 61(1), pp. 38-59, 1995.

[5] H. Blum. A Transformation for Extracting New Descriptors of Shape, In: *Models for the Perception of Speech and Visual Form*, W. Wathen-Dunn (Ed.), MIT Press, Cambridge, USA, 1967.

[6] W. Premchaiswadi, A. Tungkatsathan. *On-line Content-Based Image Retrieval System using Joint Querying and Relevance Feedback Scheme*, WSEAS Ttans. on Computers, Issue 5, Vol. 9, May 2010, pp. 465-474.

[7] K. A. Wang, H. H. Lin, P. C. Chan, C, H. Lin, S. H. Chang, Y. F. Chen, *Implementation of an Image Retrieval System Using Wavelet Decomposition and Gradient Variation*, WSEAS Ttans. on Computers, Issue 6, Vol. 7, June 2008, pp. 724-734.

[8] N. Chansiri, S. Supratid, C. Kimpan, *Image Retrieval Improvement using Fuzzy C-Means Initialized by Fixed Threshold Clustering: a Case Study Relating to a Color Histogram*, WSEAS Ttans. on Mathematics, Issue 7, Vol. 5, 2006 , pp. 926-931.

[9] H. Bakry, N. Mastorakis, *Fast Information Retrieval from Web Pages*, Proc. of the 7th WSEAS Int. Conf. on Computational Intelligence, Man-Machine Systems and Cybernetics (CIMMACS '08), 2008, pp. 229-247.

[10] A. Kulkarni, H. Gunturu, S. Datla, *Association-Based Image Retrieval*, WSEAS Trans. on Signal Processing, Issue 4, Vol. 4, April 2008, pp. 183-189.

[11] Zhe-Ming Lu, Dan-Ni Li, and Hans Burkhardt. *Image retrieval based on RST-invariant features*, IJCSNS, Vol. 6, No.2A, pp. 169 - 174 (2006)

[12] B. Reddy and B. Chatterji. *An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration*, IEEE Trans. on Image Processing 5(8), pp. 1266 - 1271 (1996)

[13] D. Zheng, J. Zhao, and A. El Saddik. *RST Invariant Digital Image Watermarking Based on Log-Polar Mapping and Phase Correlation*, IEEE Trans. on Circuits and Systems for Video Technology, Vol. XX, pp. 1-14, 2003.

[14] Z. Lu, D. Li, H. Burkhardt. *Image retrieval based on RST-invariant features*. IJCSNS, Vol. 6, No. 2A, pp. 169-174 (2006)

[15] B. Javidi (Ed.). *Image recognition and classification: algorithms, systems and applications*. Marcel Dekker Inc., NY, 2002.

[16] S. Derrode, F. Ghorbel, *Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description*, Computer Vision and Image Understanding, Elsevier Science Inc, Vol. 83, Issue 1, July 2001, pp. 57-78.

[17] J. Goodman, J. O'Rourke and P. Indyk (Eds.), *Handbook of Discrete and Computational Geometry* (2nd ed.), Ch. 39: Nearest neighbours in high-dimensional spaces. CRC Press, 2004.

[18] A. Webb. *Statistical Pattern recognition*, 2nd Ed., J. Wiley & Sons Ltd., UK (2002.)

[19] K. S. Wu and J. C. Lin, *Fast VQ Encoding by an Efficient Kick-Out Condition*, IEEE Trans.on Circuits and Systems for Video Technology, Vol. 10, No. 1, Feb. 2000, pp. 59-62.

[20] G. Qian, S. Sural, Y. Gu, S. *Pramanik, Similarity between Euclidean and cosine angle distance for nearest neighbor queries*, Proc. of the 2004 ACM symposium on applied computing, Nicosia, Cyprus, March 14-17, 2004, pp. 1232-1237.

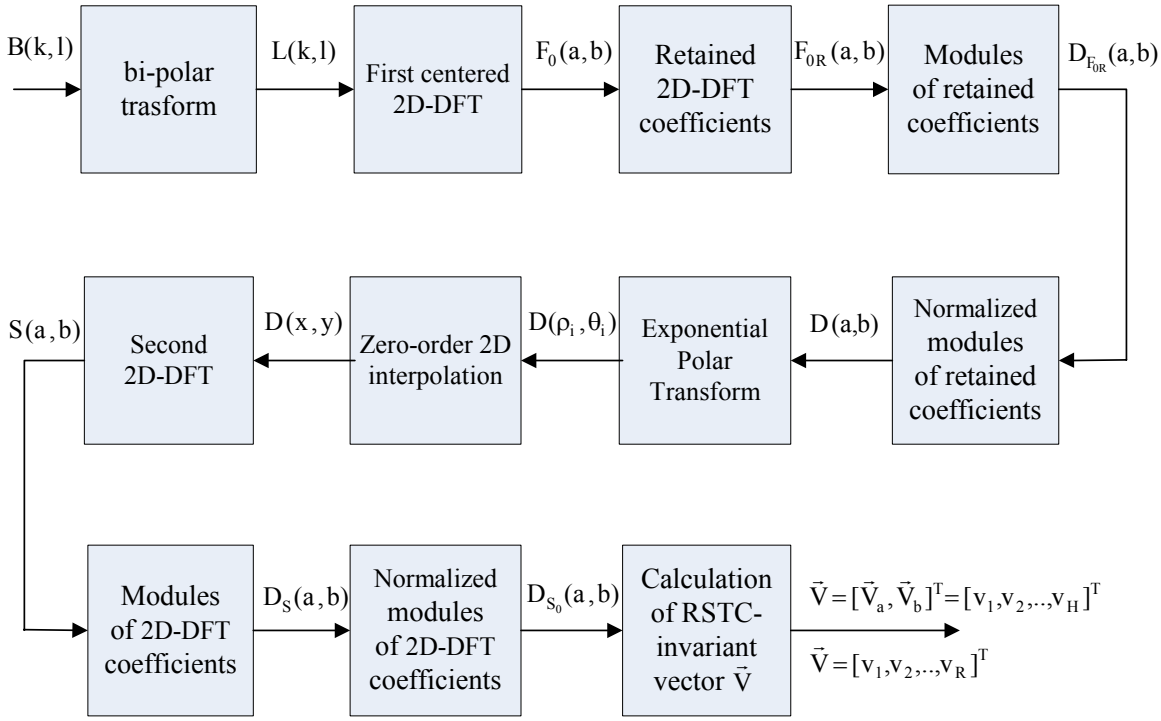


Fig. 4. Block-diagram for the calculation of H(R)-dimensional RSTC-invatiant vector, representing an object in a square window of size n×n