



Invariant pattern recognition using the RFM descriptor

Thai V. Hoang, Salvatore Tabbone

► To cite this version:

Thai V. Hoang, Salvatore Tabbone. Invariant pattern recognition using the RFM descriptor. Pattern Recognition, 2012, 45 (1), pp.271-284. 10.1016/j.patcog.2011.06.020 . inria-00614835

HAL Id: inria-00614835

<https://inria.hal.science/inria-00614835>

Submitted on 17 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Invariant Pattern Recognition Using The RFM Descriptor

Thai V. Hoang^{*,1,2} and Salvatore Tabbone²

¹*MICA Center, HUST - CNRS/UMI 2954 - Grenoble INP, Hanoi, Vietnam*

²*LORIA, CNRS/UMR 7503, Université Nancy 2, 54506 Vandœuvre-lès-Nancy, France*

Abstract

A pattern descriptor invariant to rotation, scaling, translation (RST), and robust to additive noise is proposed by using the Radon, Fourier, and Mellin transforms. The Radon transform converts the RST transformations applied on a pattern image into transformations in the radial and angular coordinates of the pattern's Radon image. These beneficial properties of the Radon transform make it an useful intermediate representation for the extraction of invariant features from pattern images for the purpose of indexing/matching. In this paper, invariance to RST is obtained by applying the 1D Fourier–Mellin and discrete Fourier transforms on the radial and angular coordinates of the pattern's Radon image respectively. The implementation of the proposed descriptor is reasonably fast and correct, based mainly on the fusion of the Radon and Fourier transforms and on a modification of the Mellin transform. Theoretical arguments validate the robustness of the proposed descriptor to additive noise and empirical evidence on both occlusion/deformation and noisy datasets shows its effectiveness.

Keywords: Invariant pattern representation; Radon transform; Fourier–Mellin transform; feature extraction; noise robustness

1 Introduction

The problem of recognition of image patterns that undergo geometric transformations like rotation, scaling, and translation is an important topic in pattern recognition and is the goal of many research works. Many approaches have been proposed for this problem and they can be classified into three main lines: brute force, image normalization, and invariant features. Brute force approaches are the most trivial ones, using “complete” training datasets; for each pattern category, the training dataset contains all its RST transformed versions. This line of approaches has inherent limitations in both storage requirement and time complexity that make them practically inapplicable. Image normalization is a solution for the reduction of the size of training datasets. The burden of the encoded RST transformation parameters in input pattern images is alleviated by normalizing them regarding their orientation, size, and position. However, despite its efficiency in the recognition stage, normalization of image patterns involves difficult inverse problems that are often ill-conditioned or ill-posed, leading to unreliable normalization results. Approaches using invariant features are based on the idea of describing image patterns by a set of measurable quantities that are insensitive to RST transformations while providing enough discriminatory power for their recognition. Mathematically speaking, if $f(x, y)$ is an image and let $g(x, y)$ be another image described as $g = \mathcal{O}(f)$, where \mathcal{O} is an RST transformation operator, then the invariant \mathcal{I} is a functional satisfying $\mathcal{I}(f) = \mathcal{I}(\mathcal{O}(f))$.

*Corresponding author. Tel: +33 3 54 95 85 60; Fax: +33 3 83 27 83 19.

Bureau B226, LORIA, Campus Scientifique, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France.

E-mail address: vanthai.hoang@loria.fr (Thai V. Hoang), tabbone@loria.fr (Salvatore Tabbone).

Many pattern descriptors have been proposed in literature for the extraction of pattern's invariant features [1, 2] using techniques that allow operator \mathcal{O} to be rotation, scaling, translation, or their combinations. Translation and scaling invariance could be obtained by using the Fourier [3] and Mellin [4] transforms respectively; rotation invariance by computing the harmonic expansion [5] or performing the discrete Fourier transform on the circular coordinate of the pattern image in polar space [6], *etc.* However, the task of combining several techniques to make operator \mathcal{O} full RST transformations while guaranteeing the discriminatory power of the invariant features is challenging and has attracted attention of many researchers. Most of the existing methods do not allow operator \mathcal{O} to be full RST transformations, they usually require normalizations for the unavailability of any of RST transformations in operator \mathcal{O} . For example, methods based on the theory of moments [7, 8] usually normalize input pattern images regarding their centroid position and size: the pattern's centroid is required to coincide with the origin of the coordinate system and the longest distance from this centroid to a pattern point is set to a fixed value. These normalizations usually introduce errors, are sensitive to noise, and thus induce inaccuracy in the later recognition/matching process.

Radon-based methods are different from the others in the sense that Radon transform is used as an intermediate representation upon which invariant features are extracted from. By applying the Radon transform on a RST transformed pattern image, the transformation parameters are encoded in the radial (for translation and scaling) and angular (for rotation) coordinates of the obtained Radon image [9]. Current techniques thus usually exploit this encoded information to define invariant features. A pioneer and notable work in this direction is the *R*-transform proposed in [10] to compute the *R*-signature. This approach uses an integral function and the discrete Fourier transform for the radial and angular coordinates of the Radon image respectively to get a 1D signature of the pattern image that is invariant to RST transformations. The strength of this approach is simplicity, however, even if an extension to a 2D signature has been proposed, the obtained signature still has low discriminatory power as there is a loss of information in the compression process from the Radon image to the 1D signature. Recently, there was an effort to apply the 2D Fourier–Mellin transform on the Radon image [11]. Similarly, Mellin transform and harmonic expansion are applied on the radial and angular coordinates of the Radon image respectively to get a pattern descriptor that is invariant to scaling and rotation. Similar efforts with different strategies to exploit the Radon image have been reported in [12–14]. These methods, however, do not allow operator \mathcal{O} to be full RST transformations as in the *R*-signature, operator \mathcal{O} could only be a combination of rotation and scaling. Moving the origin of the coordinate system to the pattern's centroid is a common solution to have translation invariance, however, this normalization step may introduce errors when the image is noisy or the pattern is slightly distorted.

Another direction in using the Radon transform for pattern description is to extract pattern features directly from the Radon image, similar to the way the Hough transform [15] is used. For example, pattern primitives in edge form are detected from the Radon image and represented analytically in [16]. Moreover, their spatial relations can be made explicit [17] and this leads to a taxonomy of pattern for its characterization [18]. This approach, however, is quite limited as it requires that the edge primitives should have analytical form. Recently, a set of spectral and structural features proposed in [19] is extracted from a slightly modified version of the Radon transform for pattern description. This set of features is not invariant to rotation and, consequently, in the matching step, these features need to be rotated to all possible angles corresponding to potential pattern's orientations in order to compute patterns' similarity. Long matching time may prevent the application of this approach in real systems. A generalization of the Radon transform, called the trace transform, has also been proposed and used for image description [20, 21] but its application is restricted due to high computational complexity.

This paper presents a novel RST invariant pattern descriptor which has briefly presented in [22],

called the RFM descriptor, defined based on the Radon, Fourier, and Mellin transforms that is invariant not only to rotation, and scaling but also to translation by definition. The proposed descriptor is a global one, its computation is essentially different from that of another type of descriptors, local descriptors [23,24], which can work only on textured patterns and fail to describe smooth or non-textured ones. The main idea here is to apply the 1D Fourier–Mellin transform, which is invariant to translation and scaling, on the radial coordinate of the pattern’s Radon image. The result of this transform is a representation of pattern that is shifted horizontally according to the pattern’s rotating angle. A further application of the discrete Fourier transform on this representation guarantees a pattern descriptor totally invariant to RST transformations. The proposed descriptor is fast in implementation and is more robust to additive noise than existing descriptors. Experimental results on both occlusion/deformation and noisy databases show the usefulness of the proposed descriptor.

The main contribution of this paper is to adopt and use the three basic transformations (Radon, Fourier, and Mellin) to transform an input pattern image into a descriptor that is totally invariant to RST transformations. Moreover, we apply the 1D Fourier–Mellin transform on the radial coordinate of the pattern’s Radon image. We use an implementation of the Mellin transform that is fast, accurate, and can preserve the discriminatory information of the input signal. Additionally, we analyzed and provided theoretical evidence for the robustness of the proposed descriptor to additive noise. And finally, we illustrate the usefulness of the proposed descriptor in the presence of occlusion/deformation and additive noise using several datasets.

The remainder of this paper is organized as follows. Section 2 gives some background on the Radon, Fourier, and Mellin transforms and their beneficial properties. Section 3 presents the combined 1D Fourier–Mellin transform and then defines the RFM descriptor. A reasonable fast and accurate implementation of the proposed descriptor is provided in Section 4 and then an analysis on the robustness of the proposed descriptor to additive noise is carried out in Section 5. Experimental results are given in Section 6 and finally conclusions are drawn in Section 7.

2 Basic material

This section reviews the three transformations used in the proposed descriptor, that are the Radon, Fourier, and Mellin transforms. It includes their definition and their derived beneficial properties.

2.1 The Radon transform

Let $f(x, y) \in \mathbb{R}^2$ be a 2D function, $L(\theta, \rho)$ be a straight line in \mathbb{R}^2 represented by

$$L = \{(x, y) \in \mathbb{R}^2 : x \cos \theta + y \sin \theta = \rho\},$$

where θ is the angle L makes with the y axis and ρ is the distance from the origin to L . The Radon transform [25] of f , denoted by \mathcal{R}_f , is a function defined on the space of lines $L(\theta, \rho)$ by the line integral along each line:

$$\mathcal{R}_f(L) = \mathcal{R}_f(\theta, \rho) = \int_L f(x, y) \, dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) \, dx dy. \quad (1)$$

In the field of shape analysis and recognition, the function $f(x, y)$ is constrained to the following particular definition:

$$f(x, y) = \begin{cases} 1 & \text{if } x \in D \\ 0 & \text{otherwise,} \end{cases}$$

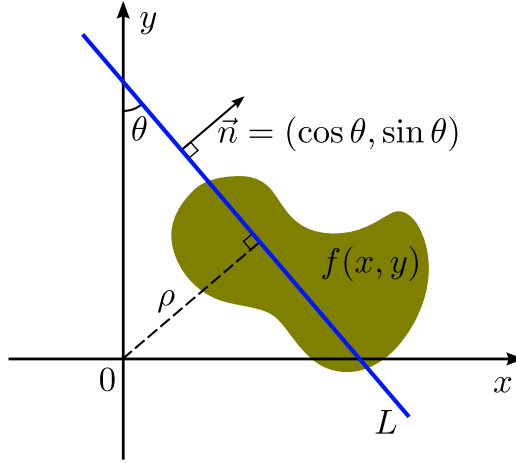


Figure 1: Graphical illustration of the Radon transform of a function $f(x, y)$. The Radon transform is a mapping from the image space (x, y) to the parameter space (θ, ρ) and can be mathematically represented by a line integral of $f(x, y)$ along all the lines L parameterized by (θ, ρ) represented in the image space (x, y) .

where D is the domain of the binary shape represented by $f(x, y)$. In the illustration of the Radon transform given in Fig. 1, the shaded region represents the region D . The value of the line integral in Eq. (1) equals the length of the intersection between the line L and the shaded region.

The Radon transform has some properties that are beneficial for pattern recognition problems as outlined below:

P1 linearity: The Radon transform is linear.

$$\mathcal{R}_{(f+g)}(\theta, \rho) = \mathcal{R}_f(\theta, \rho) + \mathcal{R}_g(\theta, \rho).$$

P2 periodicity: The Radon transform of $f(x, y)$ is periodic in the variable θ with period 2π .

$$\mathcal{R}_f(\theta, \rho) = \mathcal{R}_f(\theta + 2k\pi, \rho), \quad \forall k \in \mathbb{Z}.$$

P3 semi-symmetry: The Radon transform of $f(x, y)$ is semi-symmetric.

$$\mathcal{R}_f(\theta, \rho) = \mathcal{R}_f(\theta \pm \pi, -\rho).$$

P4 translation: A translation of $f(x, y)$ by a vector $\vec{u} = (x_0, y_0)$ results in a shift in the variable ρ of $\mathcal{R}_f(\theta, \rho)$ by a distance $d = x_0 \cos \theta + y_0 \sin \theta$, equal to the length of the projection of \vec{u} onto the line $x \cos \theta + y \sin \theta = \rho$.

$$\mathcal{R}_f(\theta, \rho) \rightarrow \mathcal{R}_f(\theta, \rho - x_0 \cos \theta - y_0 \sin \theta).$$

P5 rotation: A rotation of $f(x, y)$ by an angle θ_0 implies a circular shift in the variable θ of $\mathcal{R}_f(\theta, \rho)$ by a distance θ_0 .

$$\mathcal{R}_f(\theta, \rho) \rightarrow \mathcal{R}_f(\theta + \theta_0, \rho).$$

P6 scaling: A scaling of $f(x, y)$ by a factor α results in a scaling in the variable ρ and the amplitude of $\mathcal{R}_f(\theta, \rho)$ by the factors α and $\frac{1}{\alpha}$ respectively.

$$\mathcal{R}_f(\theta, \rho) \rightarrow \frac{1}{\alpha} \mathcal{R}_f(\theta, \alpha \rho).$$

2.2 The Fourier transform

Let $f(x) \in \mathbb{R}^1$ be a 1D function. The Fourier transform [3] of f , denoted by \mathcal{F}_f , is a function defined for every real number ξ by:

$$\mathcal{F}_f(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi\xi x} dx. \quad (2)$$

It is well known that the Fourier transform possesses a shift or translation invariant property. Consider a function $g(x) = f(x - x_0)$, a version of $f(x)$ shifted by x_0 , then:

$$\mathcal{F}_g(\xi) = \int_{-\infty}^{\infty} g(x) e^{-i2\pi\xi x} dx = e^{-i2\pi\xi x_0} \mathcal{F}_f(\xi). \quad (3)$$

Taking the magnitude of the two sides of Eq. (3) results in:

$$|\mathcal{F}_g(\xi)| = |\mathcal{F}_f(\xi)|.$$

or, in other words, the magnitude of the Fourier transform of a 1D function is invariant to translation. Discrete version of Eq. (2) defined for a sequence of numbers $\{f(n) : n = 0, \dots, N-1\}$ has the following definition:

$$\mathcal{DF}_f(k) = \sum_{n=0}^{N-1} f(n) e^{-\frac{2\pi i}{N} kn}, \quad k = 0, \dots, N-1.$$

Similarly, the discrete Fourier transform possesses a circular shift invariant property: a circular shift of the input $\{f(n)\}$ corresponds to multiplying the output $\mathcal{DF}_f(k)$ by a linear phase. Let $\{g(n)\}$ be a sequence obtained by circular shifting $\{f(n)\}$ by a distance m , then:

$$\mathcal{DF}_g(k) = e^{-\frac{2\pi i}{N} km} \mathcal{DF}_f(k). \quad (4)$$

Taking the magnitude of the two sides of Eq. (4) results in:

$$|\mathcal{DF}_g(k)| = |\mathcal{DF}_f(k)|,$$

or the magnitude of the discrete Fourier transform of a 1D function is invariant to circular shifting.

2.3 The Mellin transform

Let $f(x) \in \mathbb{R}^1$ be a 1D function. The Mellin transform [4] of f , denoted by \mathcal{M}_f , is a function defined by:

$$\mathcal{M}_f(s) = \int_0^{\infty} f(x) x^{s-1} dx, \quad (5)$$

where $s = \sigma + i\tau$. The real part of s , σ , is a constant chosen such that the integral in Eq. (5) converges. The imaginary part of s , τ , is the transform variable.

Consider a function $g(x) = f(\alpha x)$, a scaling of f by a factor α ($\alpha > 0$), then:

$$\mathcal{M}_g(s) = \int_0^{\infty} g(x) x^{s-1} dx = \alpha^{-s} \mathcal{M}_f(s). \quad (6)$$

Taking the magnitude of the two sides of Eq. (6) results in:

$$|\mathcal{M}_g(s)| = \alpha^{-\sigma} |\mathcal{M}_f(s)|.$$

Thus, except for a constant multiplicative factor $\alpha^{-\sigma}$, the Mellin transform is scale invariant. However, this scaling factor can be easily eliminated by normalization or it can be used to find the relative scale between two functions.

3 An invariant pattern descriptor based on the Radon, Fourier, and Mellin transforms

This section presents a combination of the Fourier and Mellin transforms for 1D signals and its application on the radial coordinate of the pattern's Radon image to have a pattern descriptor invariant to RST transformations. A measure of similarity for the proposed pattern descriptor is also defined, allowing the comparison/matching between any pair of images.

3.1 The combined Fourier–Mellin transform for 1D signals

Combinations of the Fourier and Mellin transforms have been proposed in literature to have signal representations that do not vary with rotation/scaling or translation/scaling. For 2D signals, they are first converted from Cartesian coordinates into polar coordinates then Fourier and Mellin transforms are performed independently on the circular and radial coordinates of the converted signals respectively [26, 27]. In this way, the obtained representation is invariant to rotation and scaling. For 1D signals, Fourier and Mellin transforms are performed in sequence directly on the signals [28, 29] to have signal representations invariant to translation and scaling.

Consider the Fourier transform of a function $g(x) = f(\alpha x - x_0)$, a scaled and translated version of $f(x)$ by parameters α and x_0 :

$$\mathcal{F}_g(\xi) = \int_{-\infty}^{\infty} f(\alpha x - x_0) e^{-i2\pi\xi x} dx. \quad (7)$$

Denoting $y = \alpha x - x_0$, Eq. (7) becomes

$$\mathcal{F}_g(\xi) = \frac{1}{\alpha} e^{-i2\pi\frac{\xi}{\alpha}x_0} \int_{-\infty}^{\infty} f(y) e^{-i2\pi\frac{\xi}{\alpha}y} dy = \frac{1}{\alpha} e^{-i2\pi\frac{\xi}{\alpha}x_0} \mathcal{F}_f\left(\frac{\xi}{\alpha}\right). \quad (8)$$

Taking the magnitude of the two sides of Eq. (8) results in:

$$|\mathcal{F}_g(\xi)| = \frac{1}{\alpha} \left| \mathcal{F}_f\left(\frac{\xi}{\alpha}\right) \right|. \quad (9)$$

The translation parameter x_0 has disappeared in Eq. (9); the remaining scaling parameter α could be eliminated by applying the Mellin transform on both sides of Eq. (9):

$$\mathcal{M}_{|\mathcal{F}_g|}(s) = \int_0^{\infty} \frac{1}{\alpha} \left| \mathcal{F}_f\left(\frac{\xi}{\alpha}\right) \right| \xi^{s-1} d\xi = \alpha^{s-1} \mathcal{M}_{|\mathcal{F}_f|}(s),$$

or:

$$|\mathcal{M}_{|\mathcal{F}_g|}(s)| = \alpha^{s-1} |\mathcal{M}_{|\mathcal{F}_f|}(s)|. \quad (10)$$

Therefore, by defining:

$$\mathcal{MF}_f(s) = |\mathcal{M}_{|\mathcal{F}_g|}(s)| = \left| \int_0^{\infty} \left| \int_{-\infty}^{\infty} f(x) e^{-i2\pi\xi x} dx \right| \xi^{s-1} d\xi \right|, \quad (11)$$

as the combined 1D Fourier–Mellin transform of a function $f(x)$ and from Eq. (10), $\mathcal{MF}_f(s)$ is invariant to translation and scaling, except for a constant multiplicative factor.

3.2 The proposed RFM descriptor

Attractive properties of the Radon and 1D Fourier–Mellin transforms lead to the proposal of a novel region-based pattern descriptor, called the RFM descriptor. The proposed descriptor of an image I is computed by $\text{RFM}(I) = \left| \frac{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(k, s)}{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(0, s)} \right|$, namely:

Step 1: The Radon transform performed on the image I , $\mathcal{R}_I(\theta, \rho)$.

Step 2: The 1D Fourier–Mellin transform performed on the radial coordinate of the obtained Radon image, $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$.

Step 3: The magnitude of discrete Fourier transform performed on the angular coordinate of the obtained, discretized Fourier–Mellin image normalized by the DC component, $\left| \frac{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(k, s)}{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(0, s)} \right|$.

Invariant properties of the proposed descriptor described above are proved as follows.

Properties. *The proposed RFM descriptor is invariant to translation, rotation, and scaling.*

Proof. Let J be the image obtained after scaling, rotating, and translating a image I using transformation parameters α , θ_0 , and $\vec{u} = (x_0, y_0)$. Properties P4–6 of the Radon transform imply:

$$\mathcal{R}_J(\theta, \rho) = \frac{1}{\alpha} \mathcal{R}_I(\theta + \theta_0, \alpha\rho - d), \quad (12)$$

where $d = x_0 \cos(\theta + \theta_0) + y_0 \sin(\theta + \theta_0)$ is the shift distance depending on θ . Eq. (12) indicates that, except for a constant multiplicative factor $\frac{1}{\alpha}$, each column constant- θ slice of the Radon image of J , $\mathcal{R}_J(\theta, \cdot)$, can be obtained by scaling and translating the constant- $\theta + \theta_0$ slice of the Radon image of I , $\mathcal{R}_I(\theta + \theta_0, \cdot)$ by a factor α and a distance d .

Applying the 1D Fourier–Mellin transform on $\mathcal{R}_J(\theta, \cdot)$ and $\mathcal{R}_I(\theta + \theta_0, \cdot)$ and using its invariant property and Eq. (10), Eq. (12) becomes:

$$\mathcal{MF}_{\mathcal{R}_J(\theta, \cdot)}(\theta, s) = \alpha^{\sigma-2} \mathcal{MF}_{\mathcal{R}_I(\theta+\theta_0, \cdot)}(\theta + \theta_0, s).$$

By varying the value of θ and s , the two images $\mathcal{MF}_{\mathcal{R}_J}(\theta, s)$ and $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$ are obtained. Moreover, except for a constant multiplicative factor $\alpha^{\sigma-2}$, $\mathcal{MF}_{\mathcal{R}_J}(\theta, s)$ can be directly obtained by shifting $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$ along the θ axis by a quantity $-\theta_0$. Applying the discrete Fourier transform on the angular coordinate of the Fourier–Mellin images $\mathcal{MF}_{\mathcal{R}_J}(\theta, s)$ and $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$ then ignoring the phase information in the coefficients, we have

$$\left| \mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_J}}(k, s) \right| = \alpha^{\sigma-2} \left| \mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(\xi, s) \right|. \quad (13)$$

This demonstrates that, except for a constant multiplicative factor $\alpha^{\sigma-2}$, the proposed descriptor computed on a scaled, rotated, and translated version J of a image I is exactly the same as the descriptor computed on the original image I . The factor $\alpha^{\sigma-2}$, however, can be easily eliminated by a normalization step using the DC component as $\left| \frac{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(k, s)}{\mathcal{DF}_{\mathcal{MF}_{\mathcal{R}_I}}(0, s)} \right|$ or it can be used to determine the relative scale between any two pattern images of the same category. \square

Calculating the proposed RFM descriptor, as described above, does not require any normalization regarding the size, position, or orientation of the patterns, it only requires a normalization in the intensity of the computed descriptor. As a consequence, the proposed descriptor is totally invariant to RST transformations.

3.3 Similarity measure

For any two images I and J represented by $\text{RFM}(I)$ and $\text{RFM}(J)$ respectively, their measure of similarity is defined as the ℓ_2 -norm distance between their descriptors as

$$\text{sim}(I, J) = \|\text{RFM}(I) - \text{RFM}(J)\|_2. \quad (14)$$

The computation of $\text{sim}(I, J)$ is simple and fast, permitting the RFM descriptor to be used in pattern matching problems with large-size datasets. More sophisticated distances like the weighted Euclidean distance [30] could be used to reduce the dominance of some of the coefficients in the RFM descriptor. However, as small value coefficients usually correspond to high frequency components, meaning that they are more sensitive to additive noise and sampling/quantization effect, balancing the coefficient contributions thus reduces the performance of the descriptor in noisy environment. Moreover, due to the orthogonality in the basis of the discrete Fourier transform, there is no correlation among coefficients of the RFM descriptor and thus the Mahalanobis distance [31], if employed, reduces to the weighted Euclidean distance.

4 Implementation of the RFM descriptor

From the definition of the RFM descriptor in Subsection 3.2, its calculation could be separated into three steps: Radon, 1D Fourier–Mellin, and discrete Fourier transforms. However, due to the interpretation of the Radon transform by means of the Fourier transform, a computational reduction is possible by fusing these two transforms. Moreover, in order to have fast implementation and high discriminatory power, a modified version of the Mellin transform has been used.

4.1 Possible fusion of Radon and Fourier transforms

The Radon transform could be computed based on recursively defined digital straight lines [32, 33] requiring $O(N \log N)$ operations for an images of $N = n \times n$ pixels. It could also be computed through 2D Fourier transform with the same complexity by means of the projection slice theorem [34], which states that the 1D constant- θ slice of the Radon image $\mathcal{R}_f(\cdot, \theta)$ and the 1D radial slice of the 2D Fourier transform $\mathcal{F}_f(\xi \cos \theta, \xi \sin \theta)$ make a 1D Fourier transform pair:

$$\begin{aligned} \mathcal{F}_{\mathcal{R}_f(\cdot, \theta)}(\xi) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(\rho - x \cos \theta - y \sin \theta) dx dy \right) e^{-i\rho\xi} d\rho \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i(x \cos \theta + y \sin \theta)\xi} dx dy \\ &= \mathcal{F}_f(\xi \cos \theta, \xi \sin \theta). \end{aligned}$$

Thus, if one chooses to implement the Radon transform through 2D Fourier transform, due to the successive application of inverse and forward Fourier transforms, performing the 1D Fourier–Mellin transform on each constant- θ slice of the Radon image is equivalent to directly performing the Mellin transform on the 1D radial slice of the 2D Fourier transform. This equivalence results in a computational reduction with no change in complexity. However, for notation convenience and for the clarity of presentation, the RFM descriptor in this paper sticks with the definition in Subsection 3.2. This choice of presentation has no impact on the remaining part. It is only when the RFM descriptor is chosen to be implemented through 2D Fourier transform that the readers should pay attention to this fusion for their own computational benefit.

4.2 The modified direct Mellin transform

The Mellin transform, as defined in Eq. (5), has a very attractive property of scaling invariance and it can be implemented optically by using an optical scale invariant correlator [26]. However, there are reported problems with its digital implementation in today's digital systems [35]. Traditionally, the Mellin transform is implemented by changing the variable $x = e^y$:

$$\mathcal{M}_f(s) = \int_0^\infty f(x)x^{s-1} dx = \int_0^\infty f(x)x^{s-1}e^{i\tau \ln x} dx = \int_{-\infty}^\infty [f(e^y)e^{\sigma y}]e^{i\tau y} dy.$$

This is, by definition, the Fourier transform of the distorted function $f(e^y)$ weighted by $e^{\sigma y}$. For sampled data, the discrete (fast) Mellin transform is implemented through FFT by re-sampling the data exponentially. Exponential sampling means interpolating the data to make them uniformly sampled in the y domain [36]. This process introduces errors into the transform and accentuates the low frequency components. Additionally, if $f(x)$ is sampled at the Nyquist rate and N is the number of data samples in the x domain then the number of the required data samples in the y domain will be $M = N \ln N$ [37]. Likewise, when $f(x)$ is nonzero at $x = 0$, $f(e^y)$ will be nonzero at $y = -\infty$ and the implementation is not realizable. In this case, the Mellin transform can be approximated by using a correction term defined based on the value of $f(x = 0)$ [37].

Another problem is on the application of the combined Fourier–Mellin transform in Subsection 3.1 for feature extraction despite it is invariant to scaling and translation. The problem is the obscurity of the discriminatory information contained in the input function [38]. Possible reasons are the discard of the phase information from the output of the Fourier and Mellin transforms and the accentuation of low frequency components. Moreover, as FFT is applied on the radial coordinate of the pattern's Radon image, the DC component is always nonzero and it, in turn, is the value of the Mellin transform's input function $f(x)$ at $x = 0$.

To avoid these problems, an alternative implementation of the Mellin transform proposed in [29], which is called the direct Mellin transform, is adopted for this work. Assuming $f(x)$ is in the form of sampled data with sampling period T , expanding Eq. (5) gives:

$$\mathcal{M}_f(s) = \int_0^T f(x)x^{s-1} dx + \int_T^{2T} f(x)x^{s-1} dx + \cdots + \int_{(N-1)T}^{NT} f(x)x^{s-1} dx. \quad (15)$$

The value of $f(x)$ is assumed to be piecewise constant in any interval T , then:

$$s\mathcal{M}_f(s) = f(0)x^s|_0^1 + f(T)x^s|_1^2 + \cdots + f((N-1)T)x^s|_{N-1}^N. \quad (16)$$

Denoting $f(iT) = f_{i+1}$, then $\Delta_k = f_k - f_{k-1}$ and, without loss of generality, assuming $T = 1$ and $f_N = 0$, Eq. (16) becomes:

$$s\mathcal{M}_f(s) = f_1x^s|_0^1 + f_2x^s|_1^2 + \cdots + f_Nx^s|_{N-1}^N = \sum_{k=1}^{N-1} k^s (f_k - f_{k+1}) = \sum_{k=1}^{N-1} k^s \Delta_k. \quad (17)$$

Substituting $s = \sigma + i\tau$ into Eq. (17):

$$(\sigma + i\tau)\mathcal{M}_f(\sigma + i\tau) = \sum_{k=1}^{N-1} k^{\sigma+i\tau} \Delta_k. \quad (18)$$

Because $k^{\sigma+i\tau} = e^{(\sigma+i\tau) \ln k}$ is bounded for any fixed constant value of σ , the right hand side of Eq. (18) is bounded, which means $|\mathcal{M}_f(s)|$ converges to zero when $|s|$ increases. This indicates the low-pass filtering of the Mellin transform.

For a set of $s_i = \sigma + i\tau_i$ with $i = 1, \dots, m$, τ_i is the arbitrary spectral component, and denoting $k^{s_i} = \phi_{i,k}$, Eq. (17) can be rewritten in matrix form as

$$\begin{bmatrix} s_1 \mathcal{M}_f(s_1) \\ s_2 \mathcal{M}_f(s_2) \\ \vdots \\ s_m \mathcal{M}_f(s_m) \end{bmatrix} = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,N-1} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m,1} & \phi_{m,2} & \cdots & \phi_{m,N-1} \end{bmatrix} \begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_{N-1} \end{bmatrix}. \quad (19)$$

The direct Mellin transform, as defined in Eq. (15), is an exact implementation of the Mellin transform for sampled data and has the following properties:

- It does not require exponential re-sampling and there is no requirement for a correction term.
- Only the differences in the value of adjacent data points are used for computing the transform.
- The coefficients $\phi_{i,k}$ in Eq. (19) can be computed off-line and stored. For each specific value of s_i , the number of stored coefficients for direct Mellin transform is N whereas the number of stored coefficients for fast Mellin transform is $N \ln N$.
- Computing $s_i \mathcal{M}_f(s_i)$ consists of only an inner product of two vectors, one of which has been pre-computed and stored and the other vector could be obtained from the input data by simple subtraction operations. Thus, Eq. (19) is very fast in digital implementation.

In order to eliminate the low-pass filtering characteristic of the direct Mellin transform, a modification is carried out by removing the $s = (\sigma + i\tau)$ factor in Eq. (18). The resulting transform is called the modified direct Mellin transform, denoted by \mathcal{MM} , as follows:

$$\mathcal{MM}_f(s) = s \mathcal{M}_f(s) = \sum_{k=1}^{N-1} k^s \Delta_k.$$

In the time domain, the modified direct Mellin transform is defined as

$$\mathcal{MM}_f(s) = s \int_0^\infty f(x) x^{s-1} dx.$$

It is easy to verify that the modified direct Mellin transform maintains the scaling invariance property and the combined Fourier-modified direct Mellin transform, defined as

$$\mathcal{MF}_f(s) = \mathcal{MM}_{|\mathcal{F}_f|}(s) = s \int_0^\infty |\mathcal{F}_f| x^{s-1} dx. \quad (20)$$

is invariant to both translation and scaling. Therefore, the 1D Fourier–Mellin transform used in the definition of the RFM descriptor in Subsection 3.2 is henceforth defined as in Eq. (20), instead of Eq. (11), and implemented through Eq. (19).

To qualitatively illustrate the invariant properties of $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$, Fig. 2 provides some example pattern images and their corresponding Radon and Fourier–Mellin images. The first row contains two original pattern images (Fig. 2(a)-2(b)) and the RST transformed versions (Fig. 2(c)-2(e)) of the pattern image in Fig. 2(b). The second row shows the Radon transform of these pattern images and the third row shows the images obtained after performing the 1D Fourier–Mellin transform on the Radon images in the second row using 150 values of τ ranging from 2.0 to 16.9 with increment of 0.1. It is noteworthy here that, due to the periodicity and semi-symmetry properties of the Radon transform (P2–3), the effective range of θ used in the computation is $0-\pi$ (rad) or $0-180$ (degree).

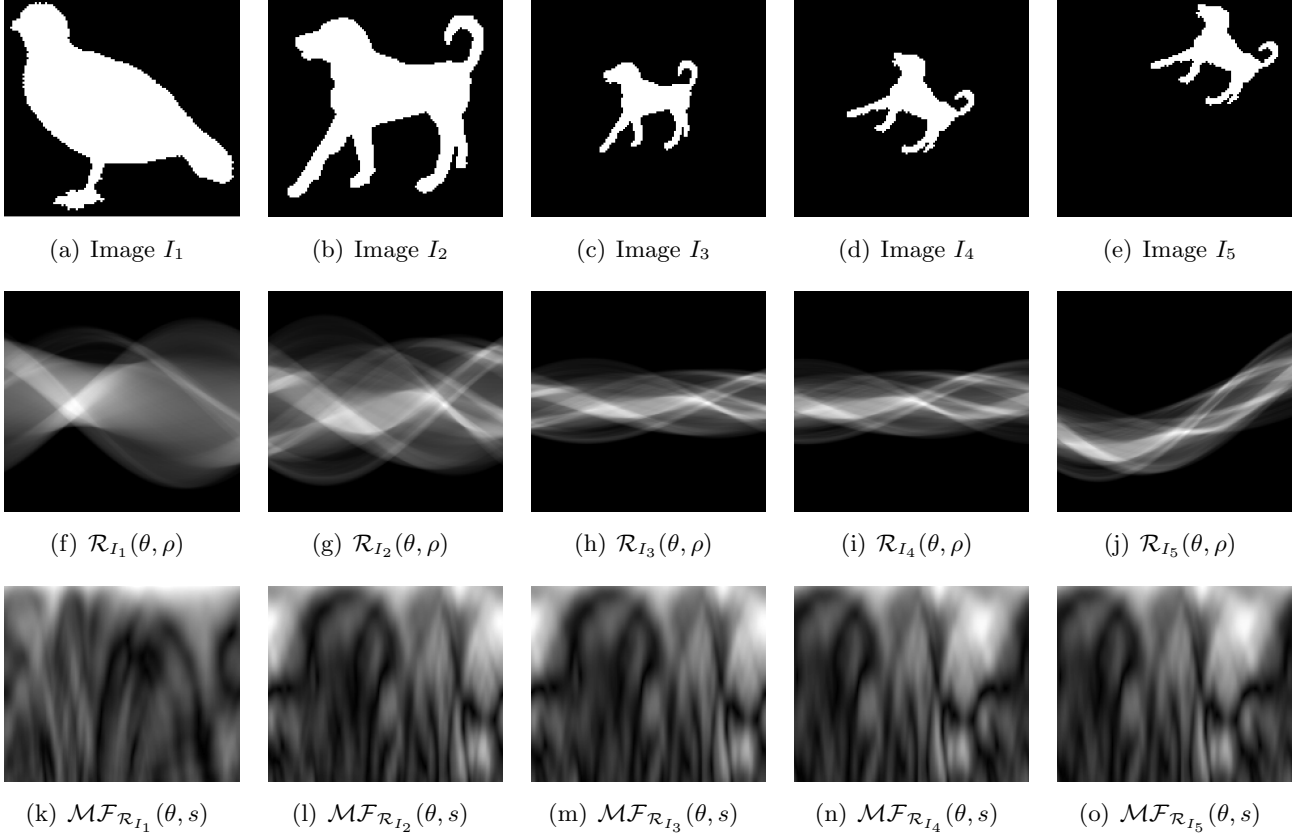


Figure 2: Illustration of the properties of the Radon and 1D Fourier–Mellin transforms performed on pattern images. The first row contains two original pattern images (a)–(b) and the RST transformed versions (c)–(e) of the pattern image in (b). The second row shows the Radon transform of these pattern images and the third row shows the images obtained after performing the 1D Fourier–Mellin transform on the Radon images in the second row using 150 values of τ ranging from 2.0 to 16.9 with increment of 0.1. The intensity of these images has been rescaled to fit the display range.

The two images I_1 , I_2 in Figs. 2(a), 2(b) are not similar and as a consequence $\mathcal{MF}_{\mathcal{R}_{I_1}}(\theta, s)$, $\mathcal{MF}_{\mathcal{R}_{I_2}}(\theta, s)$ in Figs. 2(k), 2(l) have different patterns. The images I_3 , I_4 , and I_5 in Figs. 2(c)–2(e) are transformed versions of the image I_2 in Fig. 2(b) then $\mathcal{MF}_{\mathcal{R}_{I_3}}(\theta, s)$, $\mathcal{MF}_{\mathcal{R}_{I_4}}(\theta, s)$, and $\mathcal{MF}_{\mathcal{R}_{I_5}}(\theta, s)$ in Figs. 2(m)–2(o) have the same pattern with $\mathcal{MF}_{\mathcal{R}_{I_2}}(\theta, s)$ in Fig. 2(l). The image in Figs. 2(l)–2(o) demonstrates clearly the scaling and translation invariant properties of the 1D Fourier–Mellin transform, it is invariant to scaling, translation and converts rotation in the image I into a circular shift in the variable θ of $\mathcal{MF}_{\mathcal{R}_I}(\theta, s)$.

A quantitative evaluation of the invariant properties of the Fourier–Mellin transform is given in Fig. 3 using the normalized cross correlation between all possible pairs of Fourier–Mellin images $\mathcal{MF}_{\mathcal{R}_{I_i}}(\theta, s)$ and $\mathcal{MF}_{\mathcal{R}_{I_j}}(\theta, s)$ from the third row of Fig. 2. Normalized cross correlation is selected for the sake of overcoming the constant multiplicative factor $\alpha^{\sigma-2}$ in Eq. (13). To overcome the remaining rotating parameter, the correlation is calculated for all possible rotating angle, meaning that one of the two Fourier–Mellin images can be circular shifted along its θ direction by 180 possible values from 0 to 179 with increment of 1. Denoting φ as the shifting distance, the correlation between image I_i and I_j at φ is defined as

$$C_{ij}(\varphi) = \text{corr2} \left(\mathcal{MF}_{\mathcal{R}_{I_i}}(\theta, s), \mathcal{MF}_{\mathcal{R}_{I_j}}(\theta + \varphi, s) \right),$$

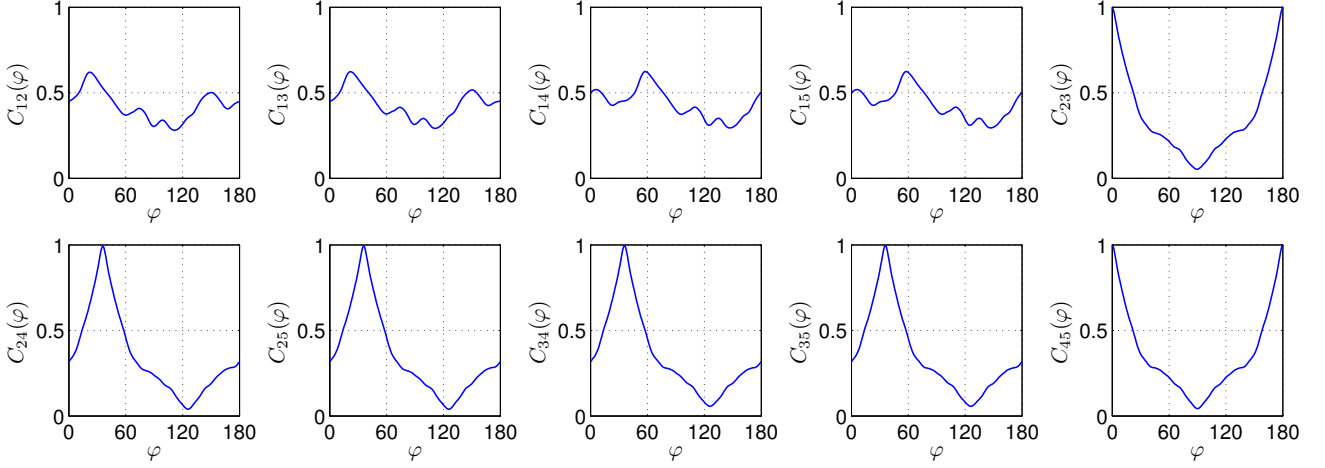


Figure 3: The normalized cross correlation between all possible pairs of Fourier–Mellin images from the third row of Fig. 2. For each pair of images, 180 correlation values are calculated by circular shifting one of the two Fourier–Mellin images along its θ direction by 180 possible values from 0 to 179 with increment of 1.

where $\text{corr2}(A, B)$ is the normalized cross correlation function between two 2D input data A and B of size $m \times n$ calculated using the following formula:

$$\frac{\sum_{i=1}^m \sum_{j=1}^n (A_{ij} - \bar{A})(B_{ij} - \bar{B})}{\sqrt{\left(\sum_{i=1}^m \sum_{j=1}^n (A_{ij} - \bar{A})^2\right) \left(\sum_{i=1}^m \sum_{j=1}^n (B_{ij} - \bar{B})^2\right)}},$$

where \bar{A} and \bar{B} are the mean values of A and B respectively.

The 10 curves $C_{ij}(\varphi)$ corresponding to the 10 possible pairs of five Fourier–Mellin images have two different patterns, $C_{ij}(\varphi)$ is peaky only when the two pattern images I_i and I_j are similar ($i, j = 2, 3, 4, 5$). The maximum value of $C_{ij}(\varphi)$ are 0.6200, 0.6236, 0.6239, 0.6239, 0.9962, 0.9943, 0.9943, 0.9970, 0.9970, 1.0000 respectively from left to right, top to bottom. The non-peaky and peaky maxima exhibit the discriminatory power of the proposed descriptor and the maximum of nearly 1 means that the Fourier–Mellin image is invariant to translation and scaling. The value of φ^* corresponding to the peak in $C_{ij}(\varphi)$ denotes the difference in orientation (in degree) between the patterns in I_i and I_j .

5 Robustness to noise

Pattern descriptor defined based on the Radon transform has the advantage of being robust to additive noise [39]. Suppose the image $f(x, y)$ is corrupted by additive white noise $\eta(x, y)$ with zero mean and variance σ^2 to be $\hat{f}(x, y) = f(x, y) + \eta(x, y)$, the Radon transform of the noisy image $\hat{f}(x, y)$ is obtained by applying the linearity property of the Radon transform (P1):

$$\mathcal{R}_{\hat{f}}(\theta, \rho) = \mathcal{R}_f(\theta, \rho) + \mathcal{R}_{\eta}(\theta, \rho).$$

Recall that the Radon transform, as illustrated in Fig. 1, is line integrals of $f(x, y)$. In the continuous domain, the Radon transform of additive white noise $\mathcal{R}_{\eta}(\theta, \rho)$ is proportional to the mean value of the noise, which means $\mathcal{R}_{\eta}(\theta, \rho) = 0$, or

$$\mathcal{R}_{\hat{f}}(\theta, \rho) = \mathcal{R}_f(\theta, \rho). \quad (21)$$

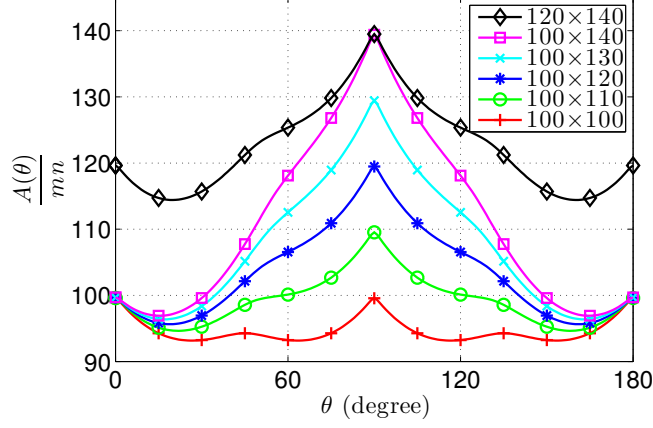


Figure 5: The values of the multiplicative factor $\frac{A(\theta)}{mn}$ in Eq. (24) at different values of θ computed on an image $f(x, y)$ of different sizes. It gets its maximum at $\theta = 90$ (degree) corresponding to the direction of the longer side of $f(x, y)$ and its minimum near $\theta = 0$ (degree). The values of $\frac{A(\theta)}{mn}$ relates directly to the noise-suppressing ability of the Radon transform.

ratio (SNR) of $\hat{f}(x, y)$ and its projection $\mathcal{R}_{\hat{f}}(\theta, \cdot)$ is:

$$\text{SNR}_{\text{image}} = \frac{\sigma_s^2 + \mu_s^2}{\sigma_n^2},$$

$$\text{SNR}_{\text{proj}(\theta)} = \frac{E_s}{E_n} = \frac{mn\sigma_s^2 + A(\theta)\mu_s^2}{mn\sigma_n^2} = \frac{\sigma_s^2 + \frac{A(\theta)}{mn}\mu_s^2}{\sigma_n^2},$$

or:

$$\text{SNR}_{\text{proj}(\theta)} = \text{SNR}_{\text{image}} + \left(\frac{A(\theta)}{mn} - 1 \right) \frac{\mu_s^2}{\sigma_n^2}. \quad (24)$$

The SNR is increased by a quantity $\left(\frac{A(\theta)}{mn} - 1 \right) \frac{\mu_s^2}{\sigma_n^2}$ after projecting $\hat{f}(x, y)$ along the direction θ . As the value of $A(\theta)$ depends both on θ , m and n , the multiplicative factor $\left(\frac{A(\theta)}{mn} - 1 \right)$ is not constant. Moreover, the value of $\frac{A(\theta)}{mn}$ is “big” because $A(\theta) = \int_{-N_\rho}^{N_\rho} n_\rho^2 d\rho$ is one order higher than $mn = \int_{-N_\rho}^{N_\rho} n_\rho d\rho$. Eq. (24) can be rewritten as

$$\text{SNR}_{\text{proj}(\theta)} \simeq \text{SNR}_{\text{image}} + \frac{A(\theta)}{mn} \times \frac{\mu_s^2}{\sigma_n^2}.$$

As $\frac{A(\theta)}{mn} \times \frac{\mu_s^2}{\sigma_n^2}$ is positive and has high value, corresponding to a high increase in the value of SNR after projection, this means that the Radon transform is very robust to additive white noise. Fig. 5 depicts the values of $\frac{A(\theta)}{mn}$ for a range of θ from 0 to 180 (degree) using input images of different sizes. We can notice that the value of $\frac{A(\theta)}{mn}$ depends on the actual size of $f(x, y)$ and it gets its maximum in the direction of the longer side and its minimum near the direction of the shorter side of $f(x, y)$, therefore:

$$\min_{\theta} \frac{A(\theta)}{mn} \simeq \min(m, n),$$

$$\max_{\theta} \frac{A(\theta)}{mn} = \max(m, n).$$

In the field of shape analysis and recognition, $f(x, y)$ is constrained to have binary values of 0 or 1 and the additive noise to $f(x, y)$ is in the form of “salt & pepper” noise, instead of white noise. To

model this type of noise, let D and d be the percentage of pixels in $\hat{f}(x, y)$ occupied by the shape region and flipped by the noise respectively, then:

$$\begin{aligned}\mu_s &= D, & \sigma_s^2 &= D - D^2, \\ \mu_n &= d(1 - 2D), & \sigma_n^2 &= d - d^2(1 - 2D)^2.\end{aligned}$$

Using Eq. (23), the signal-to-noise ratio (SNR) of $\hat{f}(x, y)$ and its projection $\mathcal{R}_{\hat{f}}(\theta, \cdot)$ is:

$$\begin{aligned}\text{SNR}_{\text{image}} &= \frac{\sigma_s^2 + \mu_s^2}{\sigma_n^2 + \mu_n^2} = \frac{D}{d}, \\ \text{SNR}_{\text{proj}(\theta)} &= \frac{mn\sigma_s^2 + A(\theta)\mu_s^2}{mn\sigma_n^2 + A(\theta)\mu_n^2} = \frac{D - D^2 + \frac{A(\theta)}{mn}D^2}{d - d^2(1 - 2D)^2 + \frac{A(\theta)}{mn}d^2(1 - 2D)^2} \\ &= \frac{D}{d} \times \frac{1 + D\left(\frac{A(\theta)}{mn} - 1\right)}{1 + d(1 - 2D)^2\left(\frac{A(\theta)}{mn} - 1\right)},\end{aligned}$$

then:

$$\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}} = \frac{1 + D\left(\frac{A(\theta)}{mn} - 1\right)}{1 + d(1 - 2D)^2\left(\frac{A(\theta)}{mn} - 1\right)}.$$

It is clear that $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ depends on the size of the input noisy image $\hat{f}(x, y)$, the projection direction θ , the percentage of shape region D , and the level of noise d . For explicit minimum value of $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$, let's say $D \in [0.3, 0.7]$ and $d \in [0, 0.2]$. This is a practically reasonable assumption as the shape usually occupies around half of the image region ($D = 0.5$) and the shape image is not too noisy (the highest level of noise used in experimentation is $d = 0.1$, meaning 10% of the image pixels is flipped, corresponding to the dataset in the bottom row of Fig. 11(b)). Due to the inverse proportion of $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ to d , $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ gets its minimum value at $\max d = 0.2$. Moreover at $d = 0.2$, as $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ decreases when D is going away from the point $D = 0.5$, the minimum value of $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$, at a specific value of $\frac{A(\theta)}{mn}$, is reached at $\min D = 0.3$. The depiction of the value of $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ for the case $\frac{A(\theta)}{mn} = 100$ over the domain $D \in [0.3, 0.7]$ and $d \in [0, 0.2]$ is given in Fig. 6(a).

Fixing $D = 0.3$ and $d = 0.2$, the dependance of $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ on $\frac{A(\theta)}{mn}$ is further given in Fig. 6(b). It is evident that $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}} > 1$, meaning the projection in the Radon transform has the property of suppressing “salt & pepper” noise. Additionally, $\frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}}$ increases with the increase in $\frac{A(\theta)}{mn}$ from 4.1667 at $\frac{A(\theta)}{mn} = 20$ (a very small image) to its limit at:

$$\lim_{\frac{A(\theta)}{mn} \rightarrow \infty} \frac{\text{SNR}_{\text{proj}(\theta)}}{\text{SNR}_{\text{image}}} = \frac{D}{d(1 - 2D)^2} = 9.375,$$

implying a better suppression of noise in the projection of bigger-size images.

6 Experimental results

In order to demonstrate the effectiveness of the proposed RFM descriptor, three experiments on grayscale and binary image datasets have been carried out. The robustness of the proposed descriptor to additive

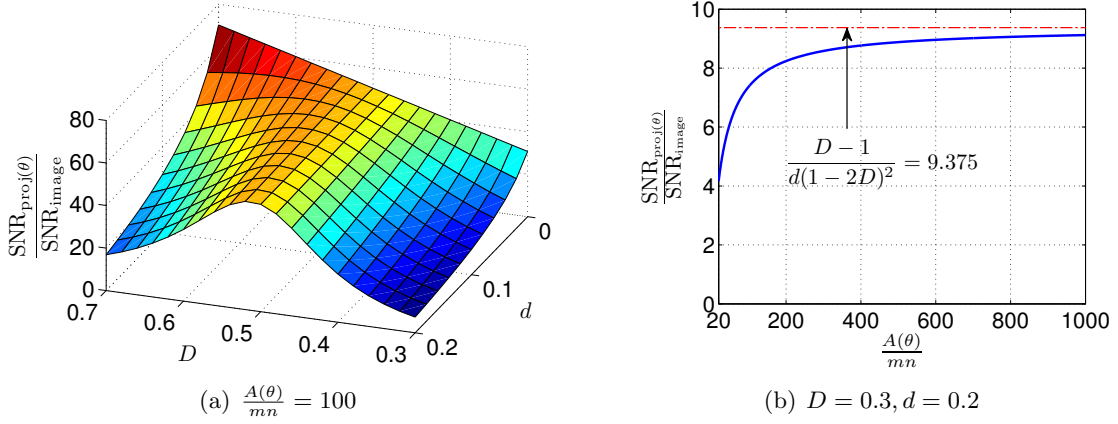


Figure 6: (a) The values of $\frac{\text{SNR}_{\text{proj}}}{\text{SNR}_{\text{image}}}$ over the domain $D \in [0.3, 0.7]$ and $d \in [0, 0.2]$ for the case $\frac{A(\theta)}{mn} = 100$, $\frac{\text{SNR}_{\text{proj}}}{\text{SNR}_{\text{image}}}$ reaches its minimum value at one of the four corners of the plotted range ($D = 0.3, d = 0.2$). (b) The dependance of the values of $\frac{\text{SNR}_{\text{proj}}}{\text{SNR}_{\text{image}}}$ on $\frac{A(\theta)}{mn}$ for the minimum case $D = 0.3, d = 0.2$.

white noise is first demonstrated by using two sets of datasets generated from images of 26 Latin characters and from the COIL-20 dataset [40] by adding white noise of different levels to them. Secondly, the proposed descriptor is computed on a set of datasets generated from the UMD Logo dataset [41] by adding “salt & pepper” noise of different levels to its images. The aim of these experiments is to demonstrate the robustness to “salt & pepper” noise of the proposed descriptor. Finally, the proposed descriptor is computed and compared with other descriptors on the reference Shapes216 dataset [42], which is used to evaluate its robustness to occlusion and deformation. Thus, the first experiment deals with grayscale patterns and the last two ones with binary patterns.

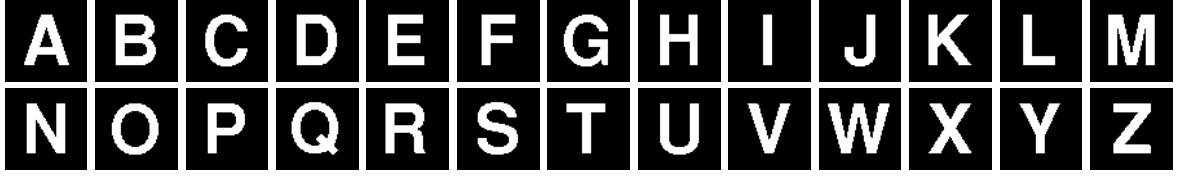
The proposed RFM descriptor is compared with angular radial transform (ART) [43], generic Fourier descriptor (GFD) [6], Zernike moments [30], R -signature [10], and Radon 2D Fourier–Mellin transforms (R2DFM) [11]. Except for the R -signature, all other comparing descriptors need normalizations in order to be invariant to RST transformations. Moreover, the R -signature and R2DFM descriptors are also defined on the Radon transform. These descriptors are selected because they are commonly used and have good reported performance. The measure used for comparison among descriptors is the precision–recall curve defined in information retrieval context.

In computing the precision–recall curve for each dataset, in the experiment, each of the images in the dataset is used as a model to which all the images in the dataset are compared/matched with. The matching is realized using the similarity measure defined in Eq. (14). The obtained matching results are then sorted, or ranked, for the determination of the n^{th} nearest matches for each model.

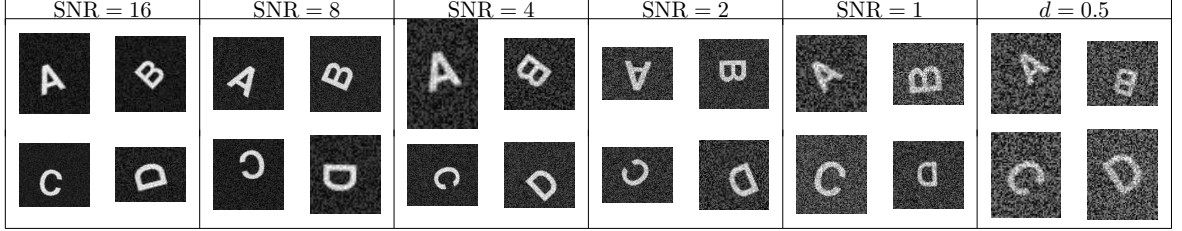
6.1 Grayscale pattern recognition

The performance of the proposed descriptor has been first tested on grayscale noisy images to demonstrate its robustness to additive white noise. Two experiments have been carried out on two different sets of datasets:

- *ExpA*: The first set of six alphabet datasets has been generated from images of 26 Latin characters as shown in Fig. 7(a). Each of these six datasets has 260 images of 26 categories, each category contains 10 images.
- *ExpB*: The second set of six object datasets has been generated from 20 object images from the



(a) Noise-free images



(b) Examples of noisy images

Figure 7: (a) Images of 26 Latin characters of size 64×64 pixels in Arial bold font used to generate the six alphabet datasets. (b) Sample images from the six alphabet datasets generated from the first four character images with six possible values of $\text{SNR} = \{16, 8, 4, 2, 1, 0.5\}$, corresponding to the six datasets.

COIL-20 dataset [40] as shown in Fig. 8(a). Each of these six datasets has 220 images of 20 categories, each category contains 11 images.

The main characteristic that differentiates *ExpA* and *ExpB*, besides the semantic content of their images, is the number of intensity levels in the original images: character images have only two levels of intensity whereas object images have multi-level intensity. Noisy images are generated from the corresponding noise-free images by randomly scaling, rotating, translating and then adding white noise to them. Let SNR be the signal-to-noise ratio defined as

$$\text{SNR} = \frac{\sum_{x,y} f^2(x,y)}{\sum_{x,y} \eta^2(x,y)}$$

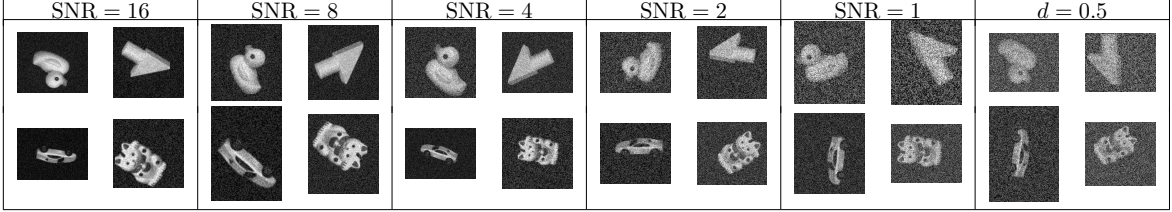
where $f(x,y)$ is the noise-free image and $\eta(x,y)$ is the added white noise, the value of SNR for each dataset is kept constant and, for each experiment, SNR has six possible values $\{0.5, 1, 2, 4, 8, 16\}$, corresponding to the six datasets. Some example images from the six datasets in the two experiments are given in Figs. 7(b), 8(b).

Comparison of the proposed RFM descriptor with ART, GFD, Zernike, *R*-signature, and R2DFM descriptors on these noisy datasets have been performed and the obtained results on *ExpA* and *ExpB* are given in Figs. 9 and 10 respectively. It is observed from these set of figures that:

- ART, GFD, Zernike, and R2DFM descriptors are not robust to additive white noise at all, their performance is similarly poor for different levels of noise.
- *R*-signature and RFM descriptors have reasonably good performance when the noise is weak ($\text{SNR} = 16, 8, 4$), demonstrating its resistance to additive noise.
- As SNR decreases (the images get noisier), the precision–recall curves of the *R*-signature and RFM descriptors generally move downwards. However, the curves of the RFM descriptor is always above that of the *R*-signature, meaning a more robustness of the RFM descriptor to noise than the *R*-signature.



(a) Noise-free images



(b) Examples of noisy images

Figure 8: (a) Twenty object images from the COIL-20 dataset [40] used to generate the six object datasets. (b) Sample images from the six object datasets generated from the four object images with six possible values of $\text{SNR} = \{16, 8, 4, 2, 1, 0.5\}$, corresponding to the six datasets.

It is thus can be concluded that the proposed RFM descriptor is more robust to additive white noise than all other comparing descriptors on grayscale noisy datasets and this provides an empirical evidence for the analytical results developed in Section 5. Poor performance of ART, GFD, Zernike, and R2DFM descriptors has its root in the required normalizations in their computation and can be explained as

- To have invariance to translation, the origin of the polar coordinate system needs to be placed at the centroid of the pattern. In the presence of noise, position of the centroid is shifted arbitrarily according to the actual noise.
- To have invariance to scaling, the radial axis is normalized by the distance from the origin of the polar coordinate system to the furthest pattern's point. In the presence of noise, this furthest point might not belong to the actual pattern but the noise.

Furthermore, it is also evident from the two set of experiments that the performance of R -signature and RFM descriptors is better on *ExpB* than on *ExpA* at each value of SNR, leading to a conclusion that the proposed descriptor performs better on multi-level than two levels grayscale images. Possible explanation for this come from the effect of re-sampling and re-quantization when an image is rotated, scaled, and/or translated. Multi-level images have “smooth” intensity surface which suffers less deformation due to re-sampling and re-quantization than the deformation occurred at sharp edge in two levels images.

6.2 Binary pattern recognition

6.2.1 Noisy datasets

The robustness of the proposed RFM descriptor to additive “salt & pepper” noise is demonstrated using six logo datasets generated from the first 25 logo images of the UMD Logo dataset as shown in Fig. 11(b). Each of these six logo datasets has 275 images of 25 categories, each category contains 11 images generated by randomly scaling, rotating, translating the original corresponding logo image and then adding “salt & pepper” noise to it. Let d be the percentage of pixels flipped by the noise, the

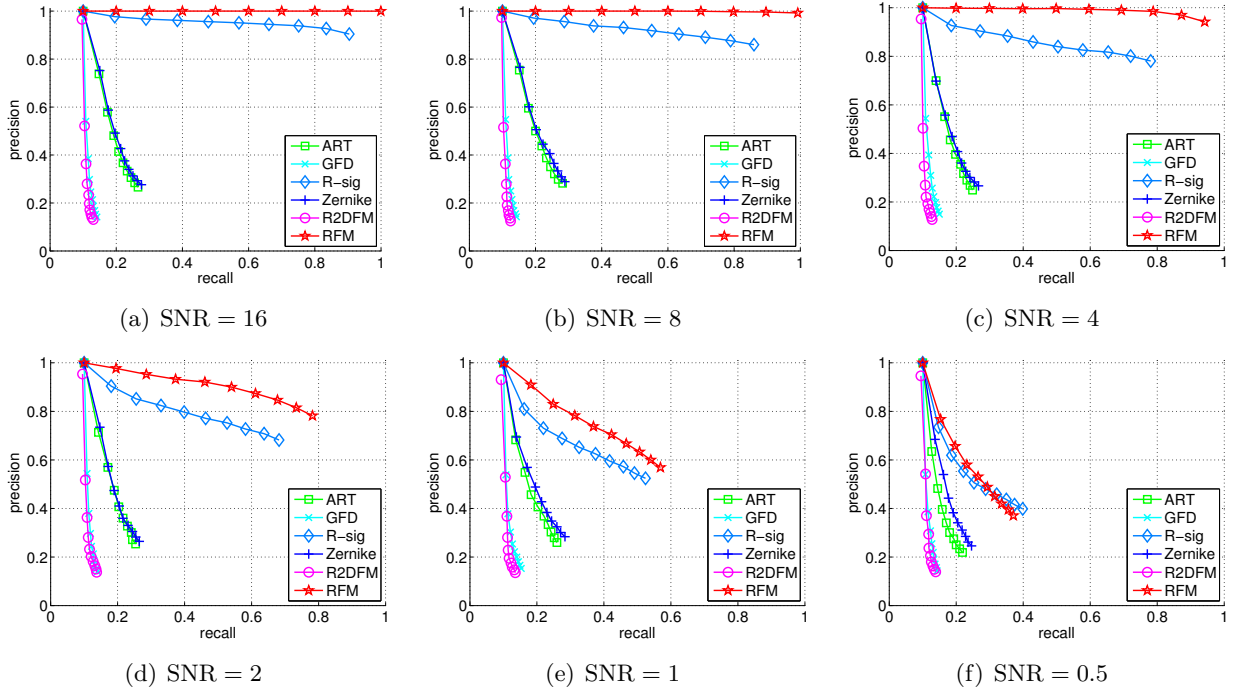


Figure 9: Precision–recall curves of comparing descriptors on the six alphabet datasets. ART, GFD, Zernike, and R2DFM descriptors are not robust to noise while *R*-signature and RMF descriptors are. As SNR decreases (the images get noisier), the precision–recall curves of the RMF and *R*-signature descriptors generally move downwards but the curves of the RMF descriptor is always above that of the *R*-signature, meaning its superiority.

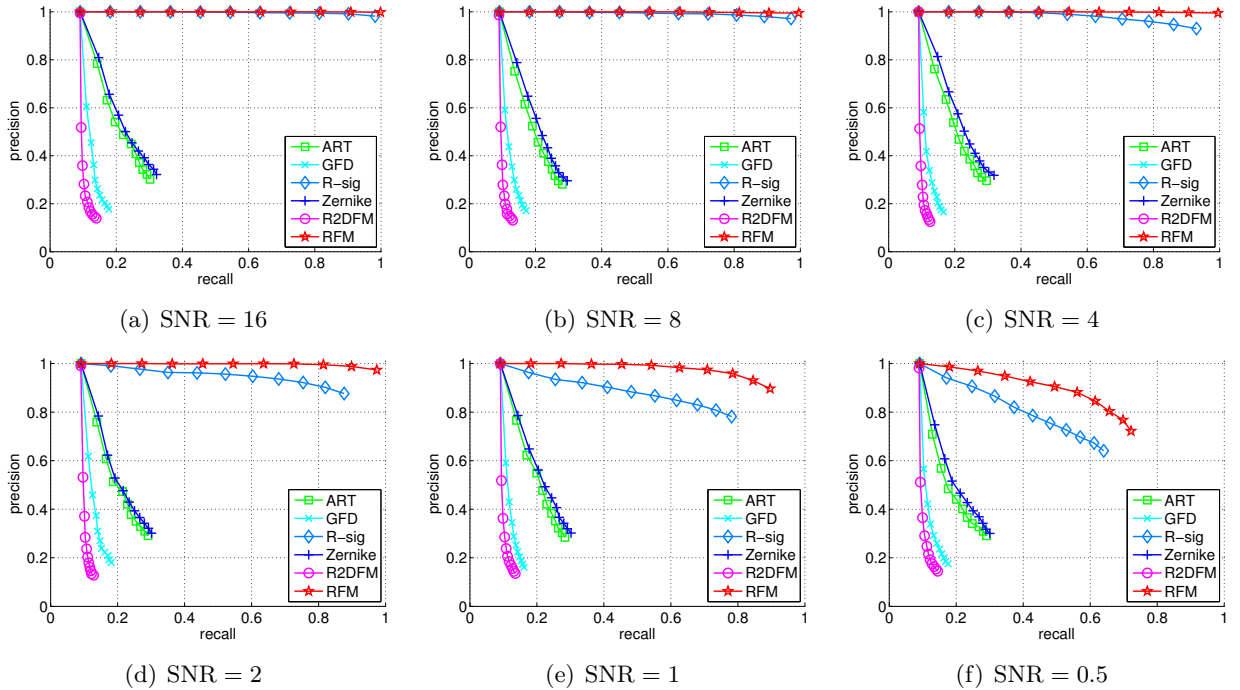
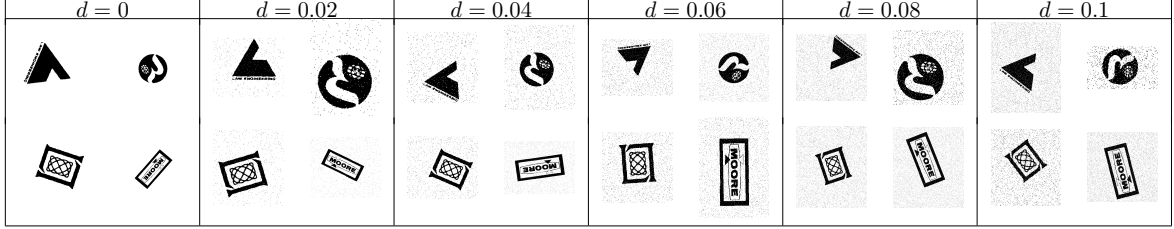


Figure 10: Precision–recall curves of comparing descriptors on the six object datasets. ART, GFD, Zernike, and R2DFM descriptors are not robust to noise while *R*-signature and RMF descriptors are. As SNR decreases (the images get noisier), the precision–recall curves of the RMF and *R*-signature descriptors generally move downwards but the curves of the RMF descriptor is always above that of the *R*-signature, meaning its superiority.



(a) Noise-free images



(b) Examples of noisy images

Figure 11: (a) Twenty-five logo images from the UMD Logo dataset [41] used to generate the six logo datasets. (b) Sample images from the six logo datasets generated from the first four logo images with six possible values of $d = \{0, 0.02, 0.04, 0.06, 0.08, 0.1\}$, corresponding to the six datasets.

value of d for each generated dataset is kept constant and d has six possible values ranging from 0 to 0.1 with increment of 0.02 corresponding to six datasets. The first dataset with $d = 0$ is actually a noiseless dataset; its use is intended for checking the invariant properties of the proposed and comparing descriptors. The values of d of the other five noisy datasets make up an arithmetic progression with common difference 0.02. These five datasets are, therefore, used to evaluate the resistance of the proposed and comparing descriptors at incrementing levels of additive noise. Some sample images from these six datasets are given in Fig. 11(b). It should also be noted that the maximum chosen value of d , 0.1, satisfies the assumption on possible value of d in the analysis of the robustness of the proposed descriptor to additive “salt & pepper” noise in Section 5.

The proposed RFM descriptor is again compared with ART, GFD, Zernike, R -signature, and R2DFM descriptors using these six datasets and the computed precision–recall curves of these descriptors are depicted in Fig. 12. For the noiseless dataset with $d = 0$ (Fig. 12(a)), all shape descriptors have ideal performance. This demonstrates that the proposed descriptor is totally invariant to both translation, rotation, and scaling. When $d \neq 0$ (Fig. 12(b)-12(f)), deterioration appears in the performance of all descriptors and their precision–recall curve moves downwards. However, the impact of d on those curves differs from one descriptor to another. Among the other descriptors, the proposed descriptor has the best performance for all the five noisy datasets while ART and Zernike descriptors have similarly worse performance. It is also observed that:

- As d increases, the curve of all the descriptors generally move downwards.
- ART and Zernike descriptors are not robust to noise at all, their performance is similarly poor for different levels of noise.
- GFD have more resistance to noise than ART and Zernike because its curves are pushed away from the ideal curve (when $d = 0$) with distance which increases along with the increase in d . However, the resistance of GFD is weaker than that of descriptors defined on the Radon transform.
- Among the three Radon-based descriptors, the shift in the curve of R -signature and RFM is more regular than that of R2DFM. In addition, and more importantly, the RFM descriptor has the

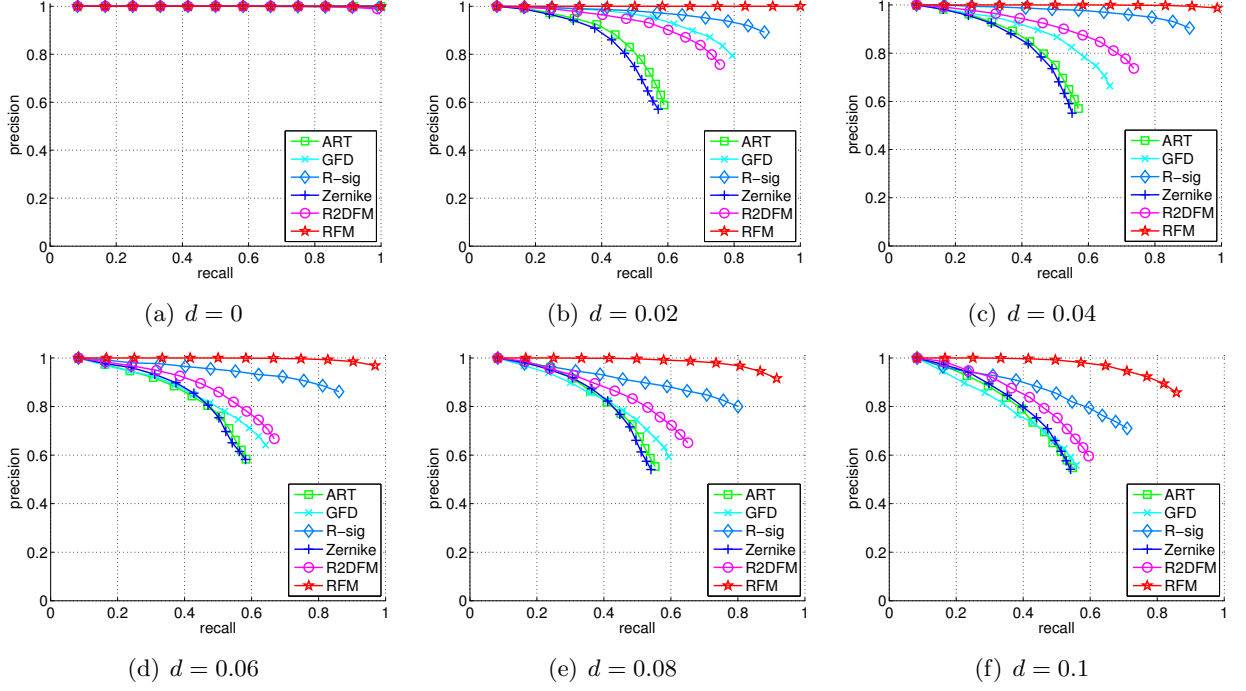


Figure 12: Precision–recall curves of comparing descriptors on the six logo datasets. For the noiseless dataset $d = 0$ (a), all shape descriptors have ideal performance. When $d \neq 0$ (b)–(f), deterioration appears in the performance of all descriptors and their curve moves downwards. However, the impact of d on those curves differs from one descriptor to another. Zernike and the RFM descriptors have worse and best performance respectively among the remaining.

smallest in the shift of its curve among all descriptors for each value of d .

The above observations lead to a conclusion that the proposed RFM descriptor is more robust to additive “salt & pepper” noise than all other comparing descriptors and this provides an empirical evidence for the analytical results developed in Section 5. Poor performance of ART, GFD, Zernike, and R2DFM descriptors in binary noisy datasets has similar explanations as in grayscale noisy datasets given in the previous subsection.

6.2.2 Occlusion and deformation dataset

Fig. 13 shows some shapes from the Shapes216 dataset. This dataset is composed of 18 shape categories with 12 samples per category, each of these cannot be obtained by RST transforming any other shape from the same category. The overall resulting nearest matches by category using the proposed RFM descriptor is given in Table 1. The value of each cell in the center region of this table represents the number of times the n^{th} nearest matches for each category are in the appropriate category. The precision of the first 12 nearest matches for each category is given in the last column and the value varies according to category. The proposed descriptor obtains:

- 100% precision in the first 12 nearest matches on *Car*, *Children*, *Face*, and *Fountain*.
- High performance on *Bone*, *Brick*, and *Classic*.
- And low performance on *Elephant*, *Fork*.

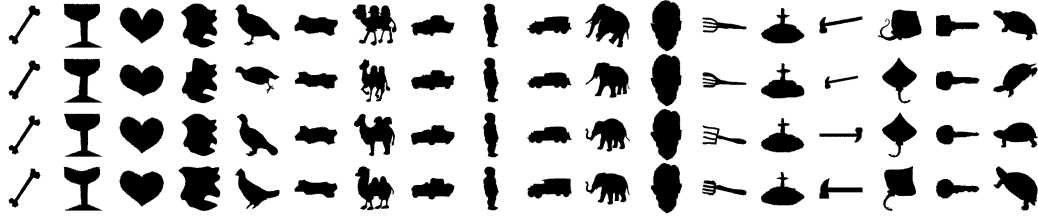


Figure 13: Example images from the Shapes216 dataset [42]. There are 18 shape categories and each category contains 12 shapes (shown in the figure are four shapes for each category), each of these cannot be obtained by RST transforming any other shape from the same category.

Table 1: The resulting nearest matches by category of the Shapes216 dataset using the proposed RFM descriptor.

Category	n^{th} nearest match												%
	1	2	3	4	5	6	7	8	9	10	11	12	
Bone	12	12	12	12	12	12	12	12	12	12	12	10	98.61
Glasse	11	12	11	12	12	11	10	11	11	12	6	12	90.97
Heart	11	8	9	8	9	5	8	9	10	9	7	7	69.44
Misk	10	10	10	6	4	10	10	10	1	10	10	10	70.14
Bird	4	1	9	10	11	6	10	10	10	11	9	7	68.06
Brick	12	12	12	12	11	12	12	12	12	12	12	12	99.30
Camel	6	6	8	9	9	5	7	5	10	8	8	9	62.50
Car	12	12	12	12	12	12	12	12	12	12	12	12	100
Children	12	12	12	12	12	12	12	12	12	12	12	12	100
Classic	11	11	11	12	12	10	11	11	12	12	11	11	93.75
Elephant	4	1	7	7	3	5	6	3	7	6	4	6	40.97
Face	12	12	12	12	12	12	12	12	12	12	12	12	100
Fork	3	2	7	4	5	1	3	5	6	5	8	8	39.58
Fountain	12	12	12	12	12	12	12	12	12	12	12	12	100
Hammer	9	11	9	11	11	8	5	8	8	8	8	10	73.61
Key	11	12	9	9	11	9	12	12	11	9	11	12	88.89
Ray	4	12	12	12	12	12	12	5	5	11	11	5	78.47
Turtle	6	9	6	6	6	5	10	6	7	4	6	5	52.78
%	100	94.44	91.67	90.74	87.04	81.02	76.85	75.93	69.44	62.96	64.35	56.94	

The last row of Table 1 provides the precision of each nearest match for all categories. It is clear that the categorization is not as good as that given in [42] where the precision of each nearest match for all categories are reported as (100, 100, 100, 100, 99, 97, 99, 96, 96, 95, 91, 80). However, it should be noted that the proposed RFM descriptor is not intended nor designed to work solely with binary patterns as in [42]. It is designed, instead, to work also with grayscale patterns under RST transformations allowing a certain level of additive noise. Furthermore, methods such as [42] cannot work with noisy patterns and need “clean” images for feature extraction, which the proposed one does not. Comparison of the proposed RFM descriptor with commonly used descriptors in this direction is given in Fig. 14. The performance of the RFM descriptor outperforms the performance of R -signature and is comparable to the performance of ART, GFD, Zernike, and R2DFM descriptors.

A better insight into the performance of proposed descriptor is by using the confusion matrix plotted as an image Fig. 14(b). The value in each table cell of row i and column j corresponds to the number of shape images of actual category j being predicted as shape images of category i . Thus, it is easy to see if the system is confusing two classes, *i.e.*, commonly mislabeling one as another. In the case of the Shapes216 dataset, it is observed that:

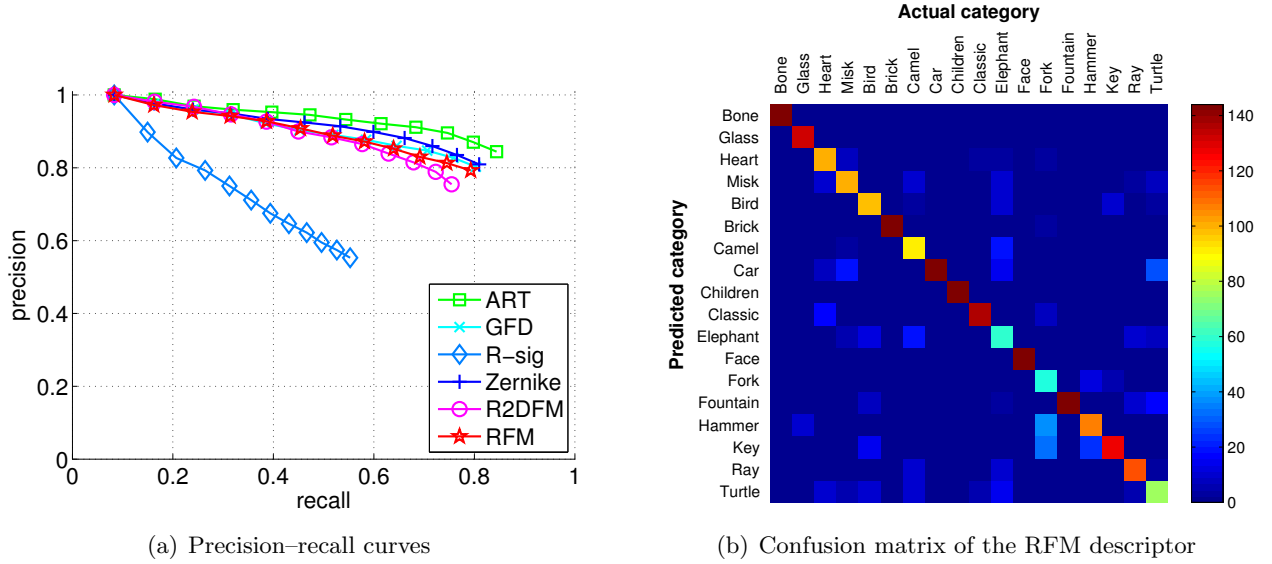


Figure 14: (a) The precision–recall curves of different descriptors on the Shapes216 dataset. The proposed RFM descriptor has similar performance with ART, GFD, Zernike, R2DFM and outperform *R*-signature. (b) The confusion matrix of the proposed RFM descriptor on the Shapes216 dataset mapped to the full range of the ‘Jet’ colormap in MATLAB (ranging from blue to red, and passing through the colors cyan, yellow, and orange).

- Low performance on *Fork* is due to its high confusion with *Hammer* and *Key*, the value in the cells of column *Fork* and rows *Hammer*, *Key* is high. This is understandable as the shapes of *Hammer* and *Key* are quite similar to the shapes of *Fork*.
- There is no high confusion between *Elephant* and any other category. However, low performance on *Elephant* is due to the diffusion of the values in the column *Elephant*. Noticeable values are in the cells of column *Elephant* and rows *Misk*, *Bird*, *Camel*, *Car*, *Ray*, and *Turtle*.

7 Conclusions

This paper presents a novel pattern descriptor defined based on the Radon, Fourier, and Mellin transforms. The proposed descriptor is obtained by applying 1D Fourier–Mellin and Fourier transforms on the radial and angular coordinates of the pattern’s Radon image respectively and has been proved to be invariant to rotation, scaling, and translation, without the need of any normalization step. The proposed descriptor’s computation is reasonably fast and correct, based mainly on the fusion of the Radon and Fourier transforms and on a modification of the Mellin transform. It is proved to be robust to additive noise both theoretically and experimentally. Experimental results show that, when compared to commonly used pattern descriptors, the proposed RFM descriptor has comparable performance on a occlusion/deformation dataset and outperforms on noisy datasets. Future work will investigate possible combinations of the RFM descriptor with local descriptors. In our perspective, due to the orthogonality between these two type of descriptors, the combination may potentially boost the performance of pattern recognition and image retrieval systems.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable and insightful comments and suggestions that help to improve the quality of the paper.

References

- [1] J. Wood, “Invariant pattern recognition: a review,” *Pattern Recognition*, vol. 29, no. 1, pp. 1–17, 1996. [1](#)
- [2] D. Zhang and G. Lu, “Review of shape representation and description techniques,” *Pattern Recognition*, vol. 37, no. 1, pp. 1–19, 2004. [1](#)
- [3] K. B. Howell, “Fourier transforms,” in *The Transforms and Applications Handbook*, A. D. Poularikas, Ed. CRC Press & IEEE Press, 2000, ch. 2. [1](#), [2.2](#)
- [4] J. Bertrand, P. Bertrand, and J. P. Ovarlez, “The Mellin transform,” in *The Transforms and Applications Handbook*, A. D. Poularikas, Ed. CRC Press & IEEE Press, 2000, ch. 11. [1](#), [2.3](#)
- [5] Y.-N. Hsu, H. H. Arsenault, and G. April, “Rotation-invariant digital pattern recognition using circular harmonic expansion,” *Appl. Opt.*, vol. 21, no. 22, pp. 4012–4015, 1982. [1](#)
- [6] D. Zhang and G. Lu, “Shape-based image retrieval using generic Fourier descriptor,” *Signal Processing: Image Communication*, vol. 17, no. 10, pp. 825–848, 2002. [1](#), [6](#)
- [7] M. R. Teague, “Image analysis via the general theory of moments,” *J. Opt. Soc. Am.*, vol. 70, no. 8, pp. 920–930, 1980. [1](#)
- [8] J. Flusser, T. Suk, and B. Zitová, *Moments and Moment Invariants in Pattern Recognition*. John Wiley & Sons, 2009. [1](#)
- [9] H. Hjouj and D. W. Kammler, “Identification of reflected, scaled, translated, and rotated objects from their Radon projections,” *IEEE Transactions on Image Processing*, vol. 17, no. 3, pp. 301–310, 2008. [1](#)
- [10] S. Tabbone, L. Wendling, and J.-P. Salmon, “A new shape descriptor defined on the Radon transform,” *Computer Vision and Image Understanding*, vol. 102, no. 1, pp. 42–51, 2006. [1](#), [6](#)
- [11] X. Wang, B. Xiao, J.-F. Ma, and X.-L. Bi, “Scaling and rotation invariant analysis approach to object recognition based on Radon and Fourier-Mellin transforms,” *Pattern Recognition*, vol. 40, no. 12, pp. 3503–3508, 2007. [1](#), [6](#)
- [12] S.-S. Xiao and Y.-X. Wu, “Rotation-invariant texture analysis using Radon and Fourier transforms,” *Journal of Physics: Conference Series*, vol. 48, no. 1, p. 1459, 2007. [1](#)
- [13] G. Chen, T. D. Bui, and A. Krzyzak, “Invariant pattern recognition using Radon, dual-tree complex wavelet and Fourier transforms,” *Pattern Recognition*, vol. 42, no. 9, pp. 2013–2019, 2009. [1](#)
- [14] X. Wang, F. xia Guo, B. Xiao, and J.-F. Ma, “Rotation invariant analysis and orientation estimation method for texture classification based on Radon transform and correlation analysis,” *J. Visual Communication and Image Representation*, vol. 21, no. 1, pp. 29–32, 2010. [1](#)
- [15] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Commun. ACM*, vol. 15, no. 1, pp. 11–15, 1972. [1](#)
- [16] V. F. Leavers and J. F. Boyce, “The Radon transform and its application to shape parametrization in machine vision,” *Image Vision Comput.*, vol. 5, no. 2, pp. 161–166, 1987. [1](#)

- [17] V. F. Leavers, "Use of the Radon transform as a method of extracting information about shape in two dimensions," *Image Vision Comput.*, vol. 10, no. 2, pp. 99–107, 1992. 1
- [18] —, "Use of the two-dimensional Radon transform to generate a taxonomy of shape for the characterization of abrasive powder particles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1411–1423, 2000. 1
- [19] Y. W. Chen and Y. Q. Chen, "Invariant description and retrieval of planar shapes using Radon composite features," *IEEE Transactions on Signal Processing*, vol. 56, no. 10-1, pp. 4762–4771, 2008. 1
- [20] A. Kadyrov and M. Petrou, "The trace transform and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 811–828, 2001. 1
- [21] M. Petrou and A. Kadyrov, "Affine invariant features from the trace transform," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 30–44, 2004. 1
- [22] T. V. Hoang and S. Tabbone, "A geometric invariant shape descriptor based on the Radon, Fourier, and Mellin transforms," in *Proc. ICPR*, 2010, pp. 2085–2088. 1
- [23] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, 2004. 1
- [24] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380–2389, 2010. 1
- [25] S. R. Deans, *The Radon Transform and Some of Its Applications*. Krieger Publishing Company, 1993. 2.1
- [26] D. Casasent and D. Psaltis, "New optical transforms for pattern recognition," *Proceedings of the IEEE*, vol. 65, no. 1, pp. 77–84, 1977. 3.1, 4.2
- [27] Y. Sheng and J. Duvernoy, "Circular Fourier-radial Mellin descriptors for pattern recognition," *J. Opt. Soc. Am. A*, vol. 3, no. 6, pp. 885–888, 1986. 3.1
- [28] R. A. Altes, "The Fourier-Mellin transform and mammalian hearing," *J. Acoust. Soc. Am.*, vol. 63, no. 1, pp. 174–183, 1978. 3.1
- [29] P. E. Zwicke and I. Kiss, "A new implementation of the Mellin transform and its application to radar classification of ships," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, no. 2, pp. 191–199, 1983. 3.1, 4.2
- [30] A. Khotanzad and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 5, pp. 489–497, 1990. 3.3, 6
- [31] P. C. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1, pp. 49–55, 1936. 3.3
- [32] W. A. Götz and H. J. Druckmüller, "A fast digital Radon transform - an efficient means for evaluating the Hough transform," *Pattern Recognition*, vol. 29, no. 4, pp. 711–718, 1996. 4.1
- [33] M. L. Brady, "A fast discrete approximation algorithm for the Radon transform," *SIAM J. Comput.*, vol. 27, no. 1, pp. 107–119, 1998. 4.1
- [34] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, and J. Waldén, "Fast Slant Stack: A Notion of Radon Transform for Data on A Cartesian Grid Which Is Rapidly Computable, Algebraically Exact, Geometrically Faithful, and Invertible," Stanford University, Tech. Rep., 2001. 4.1

- [35] S. Derrode and F. Ghorbel, “Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description,” *Computer Vision and Image Understanding*, vol. 83, no. 1, pp. 57–78, 2001. 4.2
- [36] A. De Sena and D. Rocchesso, “A fast Mellin and scale transform,” *EURASIP Journal on Advances in Signal Processing*, p. 9 pages, 2007. 4.2
- [37] G. Robbins and T. Huang, “Inverse filtering for linear shift-variant imaging systems,” *Proceedings of the IEEE*, vol. 60, no. 7, pp. 862–872, 1972. 4.2
- [38] L. H. Johnson, “The Shift and Scale Invariant Fourier-Mellin Transform for Radar Applications,” Massachusetts Institute of Technology, Tech. Rep., 1980. 4.2
- [39] K. Jafari-Khouzani and H. Soltanian-Zadeh, “Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms,” *IEEE Transactions on Image Processing*, vol. 14, no. 6, pp. 783–795, 2005. 5
- [40] S. A. Nene, S. K. Nayar, and H. Murase, “Columbia Object Image Library (COIL-20),” Department of Computer Science, Columbia University CUCS-005-96, Tech. Rep., 1996. 6, 6.1, 8
- [41] D. S. Doermann, E. Rivlin, and I. Weiss, “Applying algebraic and differential invariants for logo recognition,” *Mach. Vis. Appl.*, vol. 9, no. 2, pp. 73–86, 1996. 6, 11
- [42] T. B. Sebastian, P. N. Klein, and B. B. Kimia, “Recognition of shapes by editing their shock graphs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 550–571, 2004. 6, 13, 6.2.2
- [43] M. Bober, F. Preteux, and W.-Y. Y. Kim, “Shape descriptors,” in *Introduction to MPEG 7: Multimedia Content Description Language*, B. S. Manjunat, P. Salembier, and T. Sikora, Eds. John Wiley & Sons, 2002, pp. 231–260. 6