Rotation, scale and translation invariant pattern recognition using feature extraction

Donald Prévost a, Michel Doucet a, Alain Bergeron a, Luc Veilleux b, Paul Chevrette b and Denis Gingras a

aNational Optics Institute, 369 Franquet, Ste-Foy (Canada), G1P 4N8
bDefence Research Establishment Valcartier, 2459 Pie XI North, Courchelette (Canada), G0A 1R0

ABSTRACT

A rotation, scale and translation invariant pattern recognition technique is proposed. It is based on Fourier-Mellin Descriptors (FMD). Each FMD is taken as an independent feature of the object, and a set of those features forms a signature. FMD’s are naturally rotation invariant. Translation invariance is achieved through pre-processing. A proper normalisation of the FMD’s, gives the scale invariance property. This approach offers the double advantage of providing invariant signatures of the objects, and a dramatic reduction of the amount of data to process. The compressed invariant feature signature is next presented to a multi-layered perceptron neural network. This final step provides some robustness to the classification of the signatures, enabling good recognition behaviour under anamorphically scaled distortion. We also present an original feature extraction technique, adapted to optical calculation of the FMD’s. A prototype optical set-up was built, and experimental results are presented.

Keywords: Pattern recognition, invariance, neural networks, optical computing

1. INTRODUCTION

Invariant pattern recognition has fostered intense research activity in the last few decades. The need to achieve efficient object recognition independent of the position, scale and (in-plane) rotation is present in many practical applications. A great deal of effort was spent on correlator-based architectures. Many specialised filters 1-5 and coordinate transforms 6,7 were proposed and tested with correlator hardware. Another promising approach however is feature extraction 8-10. This method aims to describe objects by a set of numerical values associated to their morphological properties. Such a set is a compact representation of only those aspects that are interesting to the viewer (i.e. discrimination enabling), and is called a feature signature. Feature extraction is not intimately connected to a specific type of optical hardware. Depending on the desired features, the approach can be implemented via coherent or robust incoherent optical systems.

In this paper, we explore an opto-electronical means which provides real-time invariant feature signatures of objects. The definition of these features insures that they are invariant to rotation and scale changes. Translation invariance is achieved through pre-processing of the scene, with segmentation and centre of gravity calculation. Numerical simulations are presented to illustrate performance on a test set of military vehicle models. Finally, a prototype optical set-up is presented, as well as resulting experimental FMD signatures.

Further author information -
D.P.(correspondence): Email: prevost@ino.qc.ca; Telephone: 418-657-7006; Fax: 418-657-7009
M.D.: Email: doucet@ino.qc.ca
A.B.: Email: bergeron@ino.qc.ca
P.C.: Email: paul.chevrette@dre.dnd.ca
D.G.: Email: gingras@ino.qc.ca
2. FOURIER-MELLIN DESCRIPTORS

2.1. Definition

The method is based on Fourier-Mellin Descriptors (FMD). For our purpose, the FMD of order \( n,m \) will be defined as the squared modulus of the original version:

\[
H_{nm}(t) = \left| \frac{2\pi}{\pi} \int_{0}^{2\pi} f(r, \theta) e^{-i m \theta} r^n e^{i \theta} dr \right|^2, \quad n, m \in \{0, 1, \ldots \},
\]

where \( f(r, \theta) \) is the input object in polar coordinates. This definition ensures that \( H_{nm}(t) \) is real valued and is convenient for optical implementations. Every FMD of any order is inherently invariant to (in-plane) rotation of the object. Moreover, scale invariance can be achieved through a simple normalisation. Using Schwartz inequality, it can be shown that

\[
H_{n0}(t) \geq H_{nm}(t),
\]

such that:

\[
\tilde{H}_{nm}(t) = \frac{H_{nm}(t)}{H_{n0}(t)} \in [0, 1],
\]

provides normalised FMD's.

2.2. Invariant pattern recognition

The proposed scheme for invariant pattern recognition is summarised in Fig. 3. The scene is first acquired and pre-processed numerically. For this article, we assume the availability of an efficient digital segmentation technique. Pre-processing stages of this kind (segmentation/centering) can generally be implemented at (or near) video-rate, for instance on a digital systolic architecture such as Datacube™ for instance. Once an object is segmented from the scene, it is centred according to its calculated centre of gravity and fed to the FMD feature extractor, thus providing translation invariance.

Input scene \rightarrow Segmentation/Centering \rightarrow FMD signature \rightarrow Neural Network Classifier \rightarrow Classification

Fig. 1 Processing scheme of invariant pattern recognition.

The next step is to compute the signature vector \( \{ \tilde{H}_{n1}, \ldots, \tilde{H}_{nM} \} \) of \( M \) FMD's. The FMD feature extraction stage is normally computationally intensive but can be readily implemented optically. A major calculation speed-up can be expected using optics. Indeed, an optical implementation can generate the FMD features at (or near) video-rate, as will be discussed in Sec.3.

Finally, the invariant FMD signature of the input object is fed to a neural network classifier, which in turn identifies the class of the object. At this stage, this operation is efficiently achieved digitally, as the amount of data is greatly reduced to only \( M \) values, with \( M = \# \text{pixels} \) in the object plane. The classifier's generalisation capabilities are used in order to group FMD signatures, corresponding to different views of an object, into a single class. The whole process then acquires some out-of-plane rotation invariance.
3. OPTICAL FMD FEATURE EXTRACTION

3.1. Optical method

In this section, the specific task concerning the optical implementation of FMD's is addressed. By definition, FMD's involve an integral sum over complex valued functions (Eq.1). This can be expressed as the origin of the following Hankel transform:\[ H_{nm} (f) = \left[ \frac{1}{2\pi} \int_{0}^{\infty} f(r, \theta) e^{-im\theta} r \,d\theta \right]^{2} \]

which in principle lends itself well to optical implementation. However, in practice this solution suffers from some drawbacks, the main one being that the implementation of complex-valued transmittance of the form \[ g(r, \theta) = f(r, \theta) e^{-im\theta} r^{n} \]
cannot be sufficiently closely approximated on currently available spatial light modulators (SLM).

Instead, the proposed optical method is based on the decomposition of Eq.1 into a summation of real-valued terms:

\[ H_{nm} (f) = 4 \left[ \frac{C_{nm} - \frac{1}{2} C_{n0}}{2} \right]^{2} + 4 \left[ \frac{S_{nm} - \frac{1}{2} C_{n0}}{2} \right]^{2} \]

where

\[ C_{nm}(f) = \frac{1}{2\pi} \int_{0}^{2\pi} f(r, \theta) \left( \frac{1 + \cos(m\theta)}{2} \right) r^{n+1} \,d\theta \,dr \]

and

\[ S_{nm}(f) = \frac{1}{2\pi} \int_{0}^{2\pi} f(r, \theta) \left( \frac{1 + \sin(m\theta)}{2} \right) r^{n+1} \,d\theta \,dr \]

Now, as Eq.5 and 6 are integral sums of real positive functions, they can be implemented optically with amplitude modulation alone (hence on an SLM). Applying Eqs.4,5 and 6 into Eq.2 yields the following expression for normalised FMD's:

\[ R_{nm} (f) = \left[ \frac{C_{nm} - \frac{1}{2} C_{n0}}{2} \right]^{2} + \left[ \frac{S_{nm} - \frac{1}{2} C_{n0}}{2} \right]^{2} \]

The two positive real kernels of the transforms of Eq.5 and 6 are referred to as rosettes. Their mathematical expressions are respectively \[ K^{C}_{nm} (r, \theta) = \left( \frac{1 + \cos(m\theta)}{2} \right) r^{n} \] and \[ K^{S}_{nm} (r, \theta) = \left( \frac{1 + \sin(m\theta)}{2} \right) r^{n} \]. Examples of the \( K^{C}_{nm} \) and \( K^{S}_{nm} \) rosettes are illustrated on Fig.2. In order to obtain the \( C_{nm} \) (the \( S_{nm} \)) term, the image of the object \( f \) is multiplied by the \( K^{C}_{nm} \) (the \( K^{S}_{nm} \)) rosette and the resulting image is summed. The normalised FMD feature \( R_{nm} \) is calculated from Eq.7, once the \( C_{nm} \), \( S_{nm} \) and \( C_{n0} \) terms are measured. Thus, in order to get a \( M \)-feature signature of a given order \( n \), \( 2M+1 \) measurements are needed, as the zero-order term is the same for all \( M \) features.
Fig. 2  Examples of cos and sin rosettes.  Top row:  sin and cos rosettes of order \(n=0, m=5\).  Bottom row:  sin and cos rosettes of order \(n=0, m=6\).

4. SIMULATIONS RESULTS

4.1. The test-model set

Numerical simulations were carried out in order to evaluate the optical method for FMD extraction. For this purpose, a test set of four vehicle models was used, each one defining a different class: armoured personal vehicle (APV), fast armoured vehicle (FAV), tank (CHAR) and jeep (JEEP). Images of the models were taken at various combinations of viewing angle around their vertical axis (out-of-plane rotation), scale factor and in-plane rotation angle. The side view of the models is depicted on Fig. 6.

In the simulations, the first six FMD’s of order \(n=0\) \(\{R_{01}, \ldots, R_{06}\}\) make up the signature vector. The four normalised FMD signatures corresponding to the images of Fig. 3 are presented as typical examples in Fig. 4. Important inter-class variation is seen to occur so that only the first few components are necessary to properly classify the models.

4.2. Generalisation using neural networks

Normalised FMD signatures are invariant with scale changes and in-plane rotation, but vary with viewing angle (out-of-plane rotations). So, intra-class variations are likely to be important as the point of view varies. In this section, a simple hierarchical neural network architecture is used in order to perform correct classification of the signatures, considering the intra-class variations caused by out-of-plane rotations.
Fig. 3  Model set, side views. From top to bottom, left to right are the APC, FAV, CHAR and JEEP.

Fig. 4  Normalised FMD signature of the four test vehicles. Normalised FMD signatures are invariant with scale changes and in-plane rotation, but vary with viewing angle (out-of-plane rotations).
Out-of-plane rotation FMD classification

<table>
<thead>
<tr>
<th>Network parameters</th>
<th>Face-Side Net</th>
<th>Face-Id Net</th>
<th>Side-Id Net</th>
</tr>
</thead>
<tbody>
<tr>
<td># Input neurons</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td># Hidden layers</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># Neurons in hidden layers</td>
<td>10</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td># Output neurons</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td># Training set</td>
<td>21</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td># Test set</td>
<td>96</td>
<td>16</td>
<td>80</td>
</tr>
<tr>
<td># Iterations</td>
<td>112</td>
<td>28,528</td>
<td>16,328</td>
</tr>
<tr>
<td>Recognition</td>
<td>Complete</td>
<td>Complete</td>
<td>97%</td>
</tr>
<tr>
<td>Discrimination</td>
<td>n/a</td>
<td>Complete</td>
<td>97%</td>
</tr>
</tbody>
</table>

Table 1: Classification of the FMD signatures of the four test-models, for out-of-plane rotations.

Considering the nature of the models under study, a two stage hierarchical neural network was proposed. The first stage is trained to classify the signatures as to belonging to the face-view class or the side-view class. Depending on the outcome, the signature is then processed by either a specialised face-id network or a specialised side-id network, which then provides the final classification. Breaking up the classification process makes it easier to train each of the three independent networks. The networks can be made to operate in parallel, with the result from the first telling which from the face or side-id network to take the final classification from.

Experiments were performed with simulated six-components FMD’s signatures (such as those of Fig.4). They were generated from images of the models taken at a regular pace of 15-degrees out-of-plane rotation (around the vertical axis). The networks used were small, with only a few neurons and a single hidden layer. Backpropagation was used as the training technique.

The first two networks, Face-Side selection and Face-Id, were successfully trained and yielded complete discrimination over the whole range of signatures. The Side-Id network achieved some 97% discrimination, leaving an overall approximate 2% error rate. Details on the constitution and performance of the three networks are summarized in Table 1.

5. OPTICAL RESULTS

5.1. Optical set-up

A sketch of the incoherent optical set-up that was built to extract optical FMD’s is presented on Fig. 5. The set-up features two SLM’s as principal components. Its principle of operation is as follows: the centered object is presented on the surface of SLM #1, which is imaged with unit magnification onto SLM #2. The rosettes are then successively presented on SLM #2, and the total transmitted energy through both SLM’s is collected at the detector plane (using a unit detector).

Both SLM’s featured 640x480 square individually addressable graylevel pixels. The SLM’s operated at 30-frame-per-second rate, with better than 1000:1 contrast ratio. These devices were electronically addressed and operated in amplitude modulation mode (in between crossed polariser/analyser). The plane of the SLM #1 was imaged onto SLM #2 through a high-quality distortionless objective. Care had to be taken when aligning, in order to avoid Moiré effects generated by imaging a pixellised surface onto a second one. Our set-up made use of two personal computers (PC) to drive the SLM’s and a third one for the detector. This was not optimal for synchronisation purposes. In principle, a single PC equipped with the proper interfacing equipment, could be used to drive both SLM’s and to get the detector read-out.
This optical configuration is inherently sequential, requiring one video frame for every calculated term. Hence up to three frames per FMD feature were needed. Parallelism could be included however, as only a quarter of the surface was used on the SLM's to present the object or the rosette. Displaying mosaics of four replica of the object or four different rosettes, and using four spatially disjoint detectors, would speed up the calculation to but a single frame per FMD. Even more parallelism is possible, so that it is feasible to obtain a complete signature vector per video using currently available technology, for instance with a large mosaic of pre-calculated static rosette masks.

Examples of the resulting images corresponding to the optical multiplication of the object with sin and cos rosettes are presented on Fig. 6. Two of them show the APC multiplied with the sin and cos rosettes of order \([n=0, m=6]\), the other two depict the rotated CHAR, multiplied with the sin and cos rosettes of order \([n=0, m=5]\).

Some experimental FMD's calculated from optical measurements, are shown on Fig. 7. They were obtained from the set of images shown in Fig.6, only the three first FMD's are presented for each vehicle class. The stability of the experimental signatures versus in-plane rotation and scale variation was compared to the simulated results. Although experimental data shows fluctuations, good agreement is reached for three of the four models (the FAV being the worst case). Possible causes for the periodic oscillation of experimental data are a combination of remaining non-uniformities of illumination, absence of synchronisation of the devices and a slight misalignment of the axis of rotation of the objects with respect to the centre of the rosettes. Improvements on the optical FMD's stability is however desirable before out-of-plane generalisation is attempted.

Fig. 5 Sketch of the incoherent optical set-up built to extract FMD's.
A method for invariant FMD feature extraction that is suitable for optical implementation was presented. Numerical simulations were carried out on a set of four test-models. Simulated FMD signatures were generated for the four objects, with changing position, scale, in-plane rotation and point of view (vertical axis out-of-plane rotation). For any specific point of view, the signatures were invariant for the three basic invariances: translation, scale and in-plane rotation. The FMD signatures were then used as the input to a neural network classifier. Results showed that the neural network achieved a 98% correct classification rate (within the set of test-models) with changing point of view around the input object.

An incoherent optical implementation of the FMD feature extraction was achieved. This set-up provided experimental results that were shown to be similar to the numerical simulations. The opto-electronic set-up presented is extendable to parallel operation with minimal effort, opening the way to video-rate generation of invariant FMD feature signatures.

ACKNOWLEDGEMENT

This work was supported in part by contract from the Canadian Defence Research Establishment of Valcartier.
Fig. 7  FMD's of order $n=0$, $m=1,2,3$ for the four test models. From leftmost-top to rightmost-bottom: APC, CHAR, FAV and JEEP. Experimental optical data is shown by X's: $m=1$, O's: $m=2$, *'s: $m=3$ and +'s: Scale changes (0.4 to 1.0). Continuous lines are simulation results for $m=1,2,3$. 

REFERENCES