Invariant Fourier Descriptors Representation of Medieval Byzantine Neume Notation

Dimo Dimov[1], and Lasko Laskov[1,2],

[1] Institute of Information Technologies (IIT) at Bulgarian Academy of Science (BAS)
[2] Institute of Mathematics and Informatics (IMI) at BAS
Acad. G. Bonchev Str., Block 29-A, 1113 Sofia, Bulgaria
{dtdim, llaskov}@iinf.bas.bg

Abstract. During the last decade a lot of effort has been put in studying of the Fourier descriptors (FD) and their application in 2D shape representation and matching. Often FD has been preferred to other approaches (moments, wavelet descriptors) because of their properties which allow their translational, scale, rotational and contour start-point change invariance. However, there is a lack in the literature of extensive theoretical proof of these properties, which can result in inaccuracy in the methods' implementation. In this paper we propose a detailed theoretical exposition of the FDs' invariance with special attention paid to the corresponding proofs. A software demonstration has been developed with an application to the medieval Byzantine neume notation as part of our OCR system.

Keywords: Fourier descriptors, historical document image processing, OCR.

1 Introduction

Byzantine *neume* notation is a form of musical notation, used by the Orthodox Christian Church to denote music and musical forms in the sacred documents from the ancient times until nowadays. The variety and the number of different historical documents, containing neume notation is vast and they are not only a precious historical record, but also an important source of information and object of intense scientific research [10].

Naturally, most of the research of the neume notation in the historical documents is connected with the content of the documents itself, including searching for fragments or patterns of neumes, comparison between them, searching for similarities, etc. These and other technical activities are good argument in favor of creation of a software tool to help the research of the medieval neume notation. Such software tool can be an OCR (Optical Character Recognition) based system.

In the literature there are quite few attempts described for creation of a software system for processing and recognition of documents containing neume notation [5], [1]. The both works were designed to work with the contemporary neume notation in printed documents. Our goal is to develop methods and algorithms for processing and recognition of medieval manuscripts containing Byzantine neume notation with no binding to a particular notation. The main stages of the processing include: (i) preliminary processing and segmentation, [6] and [7]; (ii) symbol agglomeration in classes based on unsupervised learning of the classifier; (iii) symbol recognition.

For the goal of the unsupervised learning and recognition we need a suitable representation of the neumatic symbols which will be used for defining of a feature space which will help the comparison between the neume representatives. Since the neumes have relatively simple shapes and rarely contain cavities, in the proposed approach each neume is represented by its outer contour. For feature space definition the *Fourier transform* (FT) of the contour is used with number of high frequencies removed resulting in a reduced frequency contour representation. Such representation of 2D shapes is often called *Fourier descriptors* (FDs) [4], [8].

During the last decade FDs has been investigated in detail and applied with success in different problems, like OCR systems design [2], Content Based Image Retrieval (CBIR) [3], [4], etc. One of the main reasons FDs to be preferred to the other approaches for 2D shape representation, as moments and wavelet descriptors, are the comparatively simple methods for translation, scale, rotation and starting point normalization of FDs. This is the reason why in the literature a lot of effort is put in investigating these properties. Nevertheless, the corresponding analytical proofs are rarely given which can be the reason for inaccuracy and even errors in the implementation of the corresponding methods for linear frequency normalization of the contours.

The goal of this paper is to investigate in detail the properties of FDs to achieve their translation, scale, rotation and start- point invariance. A special attention is paid to the analytical proofs of these properties and a method for construction of linearly normalized reduced FDs (LNRFDs) for 2D shapes representation, in particular for Byzantine neume notation representation. The LNRFD representation of the neume notation can be used effectively for the goals of the unsupervised learning.

2 FD Representation of Byzantine Neume Notation

For each segmented neume symbol, the algorithm for contour finding of bi-level images [9] is applied. The resulting contour is a closed and non-self-crossing curve.

For our purposes we will represent the contour z as a sequence of Cartesian coordinates, ordered in the counterclockwise direction:

$$z \equiv \big(z(i) \mid i = 0,1,...(N-1)\big) \equiv \big(\big(x(i), y(i)\big) \mid i = 0,1,...(N-1)\big). \quad (1)$$

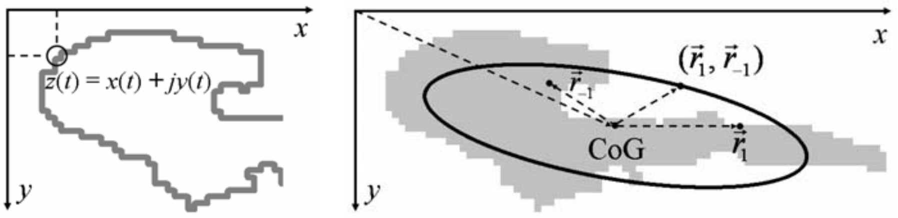
Besides, the contour is a closed curve, i.e.:

$$z(i) = z(i+N), \; i = 0,1,\ldots,N-1. \quad (2)$$

We will also assume that z is approximated with line segments between its neighboring points $z(i) = \big(x(i), y(i)\big)$, which are equally spaced, i.e.:

$$\left|z(i+1) - z(i)\right| = \left|z(N-1) - z(0)\right| = \Delta, \quad i = 0,1,\ldots,N-1, \quad (3)$$

where Δ is a constant for which we can assume $\Delta = 1$.



(a) (b)

Fig. 1. (a) Fragment of a neume contour, represented in the complex plane; (b) The contour represented as a sum of pairs of radius-vectors. The sum of the first pair gives the base ellipse of the neume symbol.

2.1 Fourier Transform of a Contour

For the sake of the FT and in correspondence with (1) we will consider the contour z as a complex function:

$$z(i) = x(i) + jy(i) = |z(i)|e^{j\varphi(i)} = |z(i)|\exp\big(j\varphi(i)\big), \quad i = 0,1,\ldots,N-1, \tag{4}$$

where x and y are its real and imaginary components in Cartesian representation, $|z|$ and $\varphi = \arg(z)$ are the respective module and phase in polar representation, and $j = \sqrt{-1}$ is the imaginary unit (see Fig.1,a). Thus, according to (2) and (3), the conditions for the DFT are fulfilled:

$$\hat{z}(k) = \mathrm{F}(z)(k) = \frac{1}{N}\sum_{i=0}^{N-1} z(i)\exp(-j\Omega ki), \ k = 0,1,\ldots,N-1, \quad \Omega = \frac{2\pi}{N}, \tag{5}$$

where $\hat{z}$ is the *spectrum* of z, $\hat{z}(k)$, $k = 0,1,\ldots,(N-1)$ are the respective *harmonics*, also called FDs, and the values $\Omega\,|\,k\,|$ have the sense of angular velocity.

The Inverse DFT relates the spectrum $\hat{z}$ to the contour z:

$$z(i) = \mathrm{F}^{-1}(z)(i) = \sum_{k=0}^{N-1}\hat{z}(k)\exp(j\Omega ki), \quad i = 0,1,\ldots,N-1, \tag{6}$$

which after equivalent transformations can be written in the form:

$$\begin{aligned} &z(i) = \text{rest} + \sum_{k=-(N_2-1)}^{N_2-1}\hat{z}(k)\exp(j\Omega ki), \quad i = 0,1,\ldots,(N_2-1), \ N_2 = \lceil N/2 \rceil, \\ &\text{rest} = \begin{cases} \hat{z}(N_2), \text{ if } N \text{ is odd} \\ 0, \quad\ \ \text{ if } N \text{ is even} \end{cases} \end{aligned} \tag{7}$$

Considered in polar coordinates, (6) and (7) lead to a useful interpretation:

Interpretation 1. The contour z represented as a sum of pairs of radius-vectors $(\vec{r}_k, \vec{r}_{-k})$, $k = 1,2...(N_2 - 1)$, $N_2 = \lceil N/2 \rceil$, rotating with the same angular velocity $\Omega|k|$, but in different directions: $\vec{r}_k$ in positive and the symmetric $\vec{r}_{-k}$ in negative direction, where $\vec{r}_k \Leftrightarrow \hat{z}(k)$ and $\vec{r}_{-k} \Leftrightarrow \hat{z}(-k) \Leftrightarrow \hat{z}(N-k)$. To this vector-sum we have also the static CoG (Center of Gravity) vector, $\vec{r}_0 \equiv \hat{z}(0)$ as well as the residual vector $\vec{r}_{N_2} \equiv \hat{z}(N_2)$ which is different from zero only if N is even (see Fig.1,b).

Apparently the terms harmonics $\hat{z}(k)$, descriptors $\hat{z}(k)$, and radius-vectors $\vec{r}_k$, $k = 0,1,...,(N-1)$ are almost identical, but express different interpretations of the contour spectrum $\hat{z}$. Thus, according to the Interpretation 1 each separate pair outlines *an ellipse* with a variable speed which direction depends on which of the two radius-vectors dominate by module. Based on this, the following practical rules can be derived:

Rule 1. The base harmonics $\hat{z}(1)$ and $\hat{z}(-1)$ cannot be zero at the same time, i.e. $|\vec{r}_1| + |\vec{r}_{-1}| \neq 0$. The opposite means that the contour is traced more than once, which is impossible with the used algorithm for contour trace.

Rule 2: If the direction of the contour trace is positive (counterclockwise), then $|\vec{r}_1| \geq |\vec{r}_{-1}|$, otherwise $|\vec{r}_1| \leq |\vec{r}_{-1}|$ (clockwise).

For concreteness we assume that the direction of the contour trace is positive, i.e. $|\hat{z}(1)| \geq |\hat{z}(-1)|$ that respects our case.

An important property of FDs is that the harmonics which correspond to the low frequencies contain the information about the more general features of the contour, while the high frequencies correspond to the details. In this sense we shall give the following definition:

Definition 1. *Reduced* FD of length L we will call the following spectral representation of the contour z :

$$\tilde{\hat{z}}(k) = \begin{cases} \hat{z}(k), & 0 \leq k \leq L \\ 0, & L < k < N_2, \quad N_2 = \lceil N/2 \rceil \end{cases} \tag{8}$$

for a boundary value L, $0 \leq L \leq \lceil N/2 \rceil$. L and respectively the frequency ΩL can be evaluated using the least-square criterion:

$$\varepsilon^2 = \frac{1}{N} \sum_{i=0}^{N-1} | z(i) - \tilde{z}(i) |^2 < \varepsilon_0^2 , \tag{9}$$

where $\tilde{z}$ is the approximation of the contour z which corresponds of the reduced frequency representation $\tilde{\hat{z}}$ and ε_0^2 is some permissible value of the criterion ε^2.

2.2 Linear Normalization of Contour in the Frequency Domain

For the aims of creation of a self-learning classifier for the neume symbols a measure of similarity between the normalized individual representatives is needed. These normalizations can be relatively easily performed in the frequency domain, using the FDs.

Translational normalization. Given (6) for the translated by a vector $\hat{z}(0)$ contour z we have that

$$z(i) - \hat{z}(0) = \sum_{k=1}^{N-1} \hat{z}(k) \exp(j\Omega ki), \quad i = 0,1,\ldots,N-1, \qquad \mathbf{(10)}$$

where $\hat{z}(0) = \frac{1}{N}\sum_{i=0}^{N-1} z(i)$ according to (5). Obviously the new contour $v \equiv z(i) - \hat{z}(0)$ coincides with the original z, but the coordinate system is translated in its CoG, i.e. the static harmonic of v is equal to zero: $\mathcal{F}(v)(0) = 0$, while all others remain unchanged: $\mathcal{F}(v)(k) = \mathcal{F}(z)(k)$, $k = 1,2,\ldots,(N-1)$.

Hence, the transitional normalization can be achieved by $z(i) := z(i) - \hat{z}(0)$, $i = 0,1,\ldots,(N-1)$, where ":=" denotes the operation assignment.

Scale normalization. Assume that we have the contour v which is a version of z, scaled by an unknown coefficient s, i.e.:

$$v(i) = sz(i), \; i = 0,1,\ldots,N-1, \; s \neq 0. \qquad \mathbf{(11)}$$

Thus, the spectral representation of v will be scaled by the same coefficient. Really, for the forward DFT of (10), it follows from (5):

$$\hat{v}(k) = \frac{1}{N}\sum_{i=0}^{N-1} sz(i)\exp(-j\Omega ki) = \frac{s}{N}\sum_{i=0}^{N-1} z(i)\exp(-j\Omega ki) = s\hat{z}(k), \; k = 0,1,\ldots,N-1 \qquad \mathbf{(12)}$$

Therefore, the scale invariance of the contour can be achieved dividing the modules of its harmonics with some non-zero linear combination of them. In the case of the algorithm of Pavlidis [9], which we use for neume contour trace, the first positive or negative harmonic is different from zero, depending on the contour trace direction. Thus, without loss of generality we may consider the module of the first harmonic is non-zero, i.e. the scale invariance can be achieved by a division of all the harmonics by it. Thus, for the spectrum $\hat{v}_s$ of the scale normalized contour v_s we have:

$$\hat{v}_s(k) = \frac{|\hat{v}(k)|}{|\hat{v}(1)|} = \frac{s|\hat{z}(k)|}{s|\hat{z}(1)|} = \frac{|\hat{z}(k)|}{|\hat{z}(1)|}, \; k = 0,1,\ldots,N-1 \qquad \mathbf{(13)}$$

Hence, scale normalization can be done by $\hat{z}(k) := \frac{|\hat{z}(k)|}{|\hat{z}(1)|}$, $k = 1,2\ldots,N-1$.

Rotational normalization. Suppose we have the contour v which is a version of the contour z, rotated to an unknown angle α. If the contours are preliminary normalized with respect to translation, i.e. their common CoG coincides with the beginning of the coordinate system, the rotation to α corresponds to multiplication of the complex representation of z with $e^{j\alpha}$.

$$v(i) = e^{j\alpha} z(i), \; i = 0,1,\ldots,N-1 \qquad \textbf{(14)}$$

The spectrum of the contour will be rotated by the same angle α. Indeed, because of the linearity of DFT, (5) and similarly to (12), for (14) we have:

$$\hat{v}(k) = \frac{1}{N}\sum_{i=0}^{N-1} e^{j\alpha} z(i)\exp(-j\Omega ki) = e^{j\alpha}\hat{z}(k), \;\; k = 0,1,\ldots,N-1 \qquad \textbf{(15)}$$

And so, the rotation by an angle α in the object domain corresponds to rotation by the same angle α of the phases of the contour spectrum.

Therefore, there are two approaches to provide the rotational invariance of the final contour representation. The *first* is to ignore the phases of the spectrum which leads to the rotationally invariant representation, but also to a big lost of information.

The *second* approach is to normalize the spectrum phases by the phase of some of the harmonics, for example the first one $\hat{v}(1)$, for which we consider again $|\hat{v}(1)| \neq 0$. Thus, for the spectrum $\hat{v}_\alpha$ of the rotationally normalized contour v_α, we have:

$$\hat{v}_\alpha(k) = \frac{\hat{v}(k)}{\exp\big(j\arg(\hat{v}(1))\big)} = \frac{e^{j\alpha}\hat{z}(k)}{e^{j\alpha}\exp\big(j\arg(\hat{z}(1))\big)} = \frac{\hat{z}(k)}{\exp\big(j\arg(\hat{z}(1))\big)}, \; k = 0,1,\ldots,N-1 \qquad \textbf{(16)}$$

Hence, rotational normalization is: $\hat{z}(k) := \dfrac{\hat{z}(k)}{\exp\big(j\arg(\hat{z}(1))\big)}$, $k = 1,2,\ldots,N-1$.

Starting point normalization. The algorithm of Pavlidis dos not guarantee that the contour trace of two identical symbols will start from one and the same start-point. The contour start-point change can be simply examined in the frequency domain.

Suppose we have the contour v which is a version of the contour z with shifted start-point by Δ positions:

$$v(i) = z(i+\Delta), \; i = 0,1,\ldots,N-1 \qquad \textbf{(17)}$$

Statement 1. Let two contours z and v, given in the complex plane, corresponds each other as (16). Then their correspondence in the frequency domain is given by:

$$\hat{v}(k) = e^{j\Omega k\Delta}\hat{z}(k), \;\; k = 0,1,\ldots,N-1. \qquad \textbf{(18)}$$

Proof: ♦ If $\Delta = 0$, then the statement is obviously true. Let us suppose that $\Delta \neq 0$. Then, for each harmonic $\hat{v}(k)$, $k = 0,1,\ldots,N-1$ from the spectrum of the contour v, according to (16) the following is true:

$$\hat{v}(k)=\frac{1}{N}\sum_{i=0}^{N-1}v(i)\exp(-j\Omega ki)=\frac{\exp(j\Omega k\Delta)}{N}\sum_{i=0}^{N-1}z(i+\Delta)\exp\big(-j\Omega k(i+\Delta)\big)$$

Using the substitution $l=i+\Delta$, we get:

$$\hat{v}(k)=\frac{\exp(j\Omega k\Delta)}{N}\sum_{l=\Delta}^{N+\Delta-1}z(l)\exp(-j\Omega kl)=\frac{\exp(j\Omega k\Delta)}{N}\left(\sum_{l=\Delta}^{N-1}z(l)\exp(-j\Omega kl)+\sum_{l=N}^{N+\Delta-1}z(l)\exp(-j\Omega kl)\right)$$

Because of the periodicity (2) of the contours $z(l)=z(l\pm N)$ we have:

$$\hat{v}(k)=\frac{\exp(j\Omega k\Delta)}{N}\left(\sum_{l=\Delta}^{N-1}z(l)\exp(-j\Omega kl)+\sum_{l-N=0}^{\Delta-1}z(l-N)\exp\big(-j\Omega k(l-N+N)\big)\right)=$$

$$=\frac{\exp(j\Omega k\Delta)}{N}\left(\sum_{l=\Delta}^{N-1}z(l)\exp(-j\Omega kl)+\sum_{m=0}^{\Delta-1}z(m)\exp(-j\Omega km)\exp(-j\Omega kN)\right)$$

But, according to (5), $\Omega N=2\pi$, and hence $\exp(-j\Omega kN)=1$. Thus, finally:

$$\hat{v}(k)=\frac{\exp(j\Omega k\Delta)}{N}\left(\sum_{l=\Delta}^{N-1}z(l)\exp(-j\Omega kl)+\sum_{m=0}^{\Delta-1}z(m)\exp(-j\Omega km)\right)=$$

$$=\frac{\exp(j\Omega k\Delta)}{N}\sum_{l=0}^{N-1}z(l)\exp(-j\Omega kl)=\exp(j\Omega k\Delta)\hat{z}(k),\ k=0,1,\ldots,N-1$$

which we had to prove ♦.

And so, according to the proved statement, the integer shift Δ of the start-point of the contour in the object domain corresponds to multiplication of the phases of its spectrum by the constant $\exp(j\Omega k\Delta)$, or equivalently to rotations of the phases as follows: the k-th phase is rotated to an angle $\delta(k)$, $\delta(k)=\Omega\Delta k,\ k=0,1,...,(N-1)$.

The invariance with respect to an arbitrary change of the contour start-point can be treated analogously to the rotational invariance, again in two approaches. The invariance in the first approach is trivial. To achieve invariance in the second approach, we propose the procedure: Normalize each harmonic of the spectrum $\hat{v}$ with the phase of the first non-zero harmonic $\hat{v}(m)\neq 0$, $\hat{v}(m)\neq 0$, as follows:

$$\hat{v}_\Delta(k)=\frac{\hat{v}(k)}{\exp\big(j\arg(\hat{v}(m))k/m\big)}=\frac{e^{j\Omega k\Delta}\hat{z}(k)}{e^{j(\Omega m\Delta)k/m}\exp\big(j\arg(\hat{z}(m))\big)}=\frac{\hat{z}(k)}{\exp\big(j\arg(\hat{z}(m))\big)},\qquad \textbf{(19)}$$
$$k=0,1,\ldots,N-1$$

Then the modified spectrum $\hat{v}_\Delta$ corresponds uniquely to the all contours that are isomorphic to the original z but with an arbitrary selected start-point.

Based on these, the start-point normalization is: $\hat{z}(k):=\dfrac{\hat{z}(k)}{\exp\big(j\arg(\hat{z}(m))k/m\big)}$, $k=m,m+1,...,N-1$, where m is the number of the first non-zero harmonic $\hat{z}(m)\neq 0$ after $\hat{z}(1)$.

The above described normalizations give us the following definition:

Definition 2. We will call *linearly normalized reduced* FD (LNRFD) of the original contour z the reduced FD of z after its processing by (10), (13), (16) and (19).

3 Conclusion

In the paper we propose an approach for constructing of LNRFDs for medieval Byzantine neume notation, which are invariant with respect to the translation, scaling, rotation and change of the contour start-point. Theoretical grounds of considered normalizations are described in more detail. For the aims of experiment, original software has been developed to extract the LNRFDs of each neume segmented in a document. These LNRFDs play the role of index into a database of neume objects.

The next stage of the proposed methodology for medieval neume notation processing and recognition will be the organization of an unsupervised learning on the basis of the above described LNRFD. After the database sorting through the LNRFD-index, the problem will be reduced to a 1D clustering problem.

Acknowledgments. This work was partially supported by following grants of IIT-BAS: Grant # DO-02-275/2008 of the National Science Fund at Bulgarian Ministry of Education & Science, and Grant # 010088/2007 of BAS.

References

1. Dalitz, C., Michalakis, G. K., Pranzas, C.: Optical recognition of psaltic byzantine chant notation. IJDAR, 11(3):143--158, (2008).
2. Dimauro, G.: Digital Transforms in Handwriting Recognition, In: Impedovo, S. (ed.): Fundamentals in HWR, NATO ASI Series "F", Vol. 124, pp.113--146, Springer-Verlag (1994).
3. Dimov, D.: Fast Image Retrieval by the Tree of Contours' Content, Cybernetics and Information Technologies, BAS, 4, 2, pp.15--29 Sofia (2004).
4. Folkers, A., Samet H.: Content-based image retrieval using Fourier descriptors on a logo database. 16th Int. Conf. on Pattern Recogn., 3:521--524, (2002).
5. Gezerlis, V., Theodoridis, S.: Optical character recognition of the orthodox hellenic byzantine music notation. Pattern Recognition, 35(4):895--914, (2002).
6. Laskov, L., Dimov, D.: Color Image Segmentation for Neume Note Recognition, In: Proceedings of the Int.Conf. A&I'07, 03-06.10.07, pp.III.37--41, Sofia (2007).
7. Laskov, L., Dimov, D.: Segmentation of Ancient Neumatic Musical Notation, In: Proceedings of the Int.Conf. A&I'08, 01-04.10.08, pp.II.21--24, Sofia (2008).
8. Zhang, D., Lu, G.: A comparative study on shape retrieval using Fourier descriptors with different shape signatures. In Intelligent Multimedia, Computing and Communications Technologies and Applications of the Future, pp1--9, Fargo, North Dakota, USA, June (2001).
9. T. Pavlidis. Algorithms for Graphics and Image Processing. Springer (1982).
10. Distributed Digital Library of Chant Manuscript Images, http://www.scribeserver.com/NEUMES/distributed_image_library/view/about.htm