

2-D FORM DESCRIPTORS BASED ON A NORMALIZED PARAMETRIC POLAR TRANSFORM (UNL TRANSFORM)

T. W. Rauber and A. S. Steiger-Garção

Universidade Nova de Lisboa – Faculdade de Ciências e Tecnologia
Departamento de Informática – Intelligent Robotics Group
2825 Monte de Caparica – Portugal
Tel. +351-1-2953220 – Fax +351-1-2955641 – E-Mail tr@fct.unl.pt

ABSTRACT

An essential research objective in artificial vision are shape descriptors which are invariant for translation, scale changes and rotations of a bidimensional pattern. A variety of approaches has proved the capacity to characterize forms, like signatures, 1-D Fourier descriptors, moment invariants, Complex-log (Log-polar) transform or Fourier transform. None of these techniques can claim a general purpose applicability to every kind of 2-D pattern.

In this paper a method is proposed for the shape description of arbitrary complex forms that are composed by parametric curves. This constraint of representability is satisfied for most patterns. Regions with holes e.g. can equivalently be characterized by the contours of the significant parts (boundaries, holes).

A method, denoted as *UNL transform* performs a normalization operation for translation and scale changes and causes rotations to appear as periodic translations in the transformed representations of the pattern. It creates an optimal input for a 2-D Fourier image transform which yields the numerical descriptors called *UNL Fourier Features*. It permits the mapping of any shape to a single vector (or point in a n -dimensional space).

The distinctive character of the approach will be pointed out, especially to the Complex-log transform.

The analytic theory of the UNL transform is introduced, together with practical concepts to apply it in the discrete case where a pattern is given as a bitmap. Finally experimental results for a classification task are presented together with conceptual limitations of the approach.

INTRODUCTION

Bidimensional black and white shapes are either characterized by their contours and/or by their region if they possess an area. The digit "5" is an example for the first category of patterns and a box with a hole inside for the second category. This most general classification [1] determines if a certain form descriptor method can be applied for the characterization of the pattern or not. In order to evaluate the improvement that our method represents compared with already existing methods, a benchmark pattern is proposed. The stylized head of a robot in fig. 1 will be used for this purpose. Parts of it could be modelled by a region-based technique, like the head with eyes, nose and mouth. The antenna however are only representable by curves.

For a general overview about shape representation methods consult [1], [2], [4], [5].

Signatures [1], [3], [6] are only applicable locally to parts of the test pattern. The eyes, nose, mouth and boundaries of the head can be represented as signatures. The antenna and most important the figure as a whole are not mappable to a single signature. Besides concave boundaries are an insurmountable limit for this approach. That restriction of only local applicability is also valid for Chain codes [1], [7] and polygonal approximations [1], [8], [9]. Fourier descriptors (FD) [1], [10], [11] have similar limitations as signatures. For each primitive form that composes the test pattern a proper FD is necessary. No global Fourier series is available to store the form of the pattern in its totality. A further drawback of FDs are that the patterns in general are limited to closed curves.



Fig. 1 Test pattern.

Moments (normalized central moments) are used to describe 2-D images [1], [12]. They are invariant to TSR transformations (translations, scale and rotations). A representative application of moments together with FDs to recognize airplane silhouettes can be found in [13]. See also [14],[15],[16]. Like the 2-D Fourier transform and the Complex-Log transform which will be characterized below, moments are image oriented features. Basically one can state that the input for moment invariants are pixel values sampled at a certain image point (x,y) . Moment invariants are not generated by analytic information about the object geometry, but rather by an *instantiation* of an object on a 2-D pixel matrix. The shape transformations (TSR) are tried to be compensated for, regarding the pixel values. For our test pattern the method of moment invariants is only partially suited because the pattern consists only of fine lines. In all experiments that were based on moment invariants the patterns had an interior region. This suggests the empirical conclusion that moments are not appropriate to describe patterns where only boundary information is available.

Special attention must be paid to the Complex-Log transform (Log-polar transform) because it has some characteristics in common with the UNL transform. The basis for this form descriptor is the simulation of mapping structures commonly found in the visual system of human beings [17],[18],[19]. In [20] the Log-polar transform was successfully applied for 2-D shape recognition. A further analysis was performed in [21], where especially the drawbacks were mentioned. For an application in character recognition see [22]. This technique will be directly compared to the UNL transform in the next section.

The 2-D Fourier transform (FT) of an image is a valuable tool for the analysis of objects inside a scene. The FT is also pixel oriented, like moments and the Complex-Log transform. The input for the FT is a pixel value sampled from an image coordinate: $f(x,y)$. For an introduction to the FT see e.g. [1] and [2]. Its major drawback for direct pattern recognition is its rotation variance. If only translations of objects in a scene occur the magnitudes of the FT are invariant. The UNL transform will produce an image in which at most periodic translations appear on one axis. Therefore we will use the FT to produce the final shape descriptors, the UNL Fourier Features. For our test pattern the FT applied directly to the original image is not an appropriate shape descriptor.

In order to complete the overview of existing 2-D shape descriptors one should also mention syntactic techniques [23] and autoregressive models [24],[25].

THE UNL TRANSFORM

In this section we will draw the mathematical framework for our 2-D pattern descriptor. What distinguishes the proposed method from other shape representation schemata, especially the Complex-Log transform? The most important attribute of the UNL transform is its analytic approach to perform a pattern transformation. It implements a coordinate transform for parametric curves from Cartesian to normalized polar coordinates. The analytic equations of the pattern curves must be known to be able to transform the pattern. If these equations are not known a priori which in practice is mostly the case, they must be estimated. This estimation is based on a binary pixel image.

Definition

A formalization is given for the UNL transform and the UNL Fourier Features. Mathematical details will be shifted to the respective appendices. The notion of complex numbers will be used to represent 2-D coordinates. This allows a concise formalism. Hence a Cartesian point (x,y) is represented by a complex number $z = x + jy$.

Definition 1: Let an object Ω be composed by a finite set of smooth parametric curves $z(t)$ in the Cartesian coordinate system:

EQ1
let $O = (O_x, O_y)$ be the centroid of all curves (appendix A)

$$\Omega(t) = \bigcup_{i=1}^n z_i(t) \quad z_i(t) = x_i(t) + jy_i(t) \quad t \in (0, 1)$$

and let M be the maximum Euclidean distance from O to all curve points:

$$M = \max_{i,t} \{ \|z_i(t) - O\| \} \quad \forall i = 1 \dots n \quad t \in (0, 1)$$

EQ2

Then a coordinate transform U for each curve $z(t)$ exists which represents the object in a normalized polar coordinate system with origin O .

EQ3

$$U: ((0, 1) \rightarrow C) \rightarrow ((0, 1) \rightarrow C)$$

$$U(z(t)) = R(t) + j \times \theta(t)$$

$$= \frac{\|z(t) - O\|}{M} + j \times \text{atan} \left(\frac{y(t) - O_y}{x(t) - O_x} \right)$$

Lemma 1 (Proof in Appendix B): The transformed object $U(\Omega)$ is invariant to translations of Ω by the offset Δz and invariant to scale changes by a scalar a . A rotation about any point z_r by the angle $\Delta\theta$ has no effect on R and causes a 2π -periodic translation of θ by $\Delta\theta$ (cyclic shift).

N.B.: The operator U maps curves to curves by a coordinate transform ($U: (\text{Curve}) \rightarrow (\text{Curve})$). It does not map a single point from the Cartesian plane to the polar plane.

Definition 2: Let $U(\Omega)$ be the UNL transformation of object Ω , let the pointset I be the *image (trace)* of $U(\Omega)$, i.e. the image of all parametric curves that compose $U(\Omega)$ (appendix C). Then the magnitudes of the 2-D Fourier transform of I are the *UNL Fourier Features (UFF)*.

Theorem 1: The UNL Fourier Features are invariant to translations, scale changes and rotations of the original object Ω .

Proof: This is true following the translation theorem of the Fourier transform and Lemma 1.

Example of an UNL Transform of a Curve Pattern

For the test pattern of fig. 1 the analytic UNL transform is presented. The calculus of the transformed pattern was analytically performed by *Mathematica* [26]. The centroid falls inside the nose. The maximum distance from the centroid appears two times at the end of the two antenna. The three "blobs" in the middle are the transformed eyes and mouth. The four peaks are the head and the two spikes are the transformed antenna. Corresponding points are labeled with the same character in both the original and transformed pattern. As an example consider point f . The vector from the centroid to f has an angle of 45 degrees ($\pi/4$) to the horizontal reference axis and has a distance to the centroid of about 75% in relation to the maximum distance at point a or b .

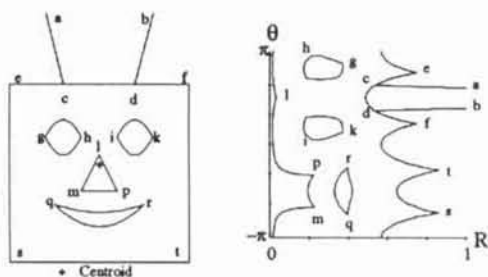


Fig. 2 UNL Transform of the test pattern.

Comparison UNL Transform – Log-Polar Transform

The differences between the UNL transform (U) and the Log-Polar transform (Complex-Log transform) (L) must be explicitly stated because both techniques have some approaches in common.

Both seek a translation normalization by shifting the centroid of the object pattern to the origin of a polar coordinate system. Both map from originally Cartesian coordinates to polar coordinates.

The fundamental differences however are:

Application Constraints: While U can be applied to every pattern composed of curves, L is limited to patterns which possess a region. For the test pattern of fig. 1 L is not suited, e.g. it could not meaningfully be used to transform the antenna.

Subject of Mapping: L maps image points through a mathematical function to other image points. A complex coordinate z is transformed to $\ln z$. The axes of the original (x,y) -pattern change to $(\ln \|z\|, \theta)$. U is "just" a coordinate transform. It does not change the value of z but rather represents it in another coordinate system.

Mathematical Operations: L calculates $\ln z$, which yields another complex number. U is mathematically an identity operation. The value of z is invariant.

Scale invariance: U is analytically scale invariant for every positive scale factor. The limitations only occur in the discrete pixel matrix when the scale factor becomes too big. L is only obviously scale invariant. In [21] the problem of *stretching* is highlighted. Since the scaled coordinate $a*z$ is transformed to $\ln z + \ln a$ [20], the transformed pattern not only dislocates its pixels but also distorts them. This defect only permits the application of L under moderate scaling conditions. Particular problems occur when the centroid's coordinates of the pattern fall inside the area of an object's hole. In this case it becomes very hard to match the pattern with the prototypes without using heuristics. In [20] "useless pixels" are cut and in [21] it is tried to compensate for the stretching with a Laplacian filter.

CALCULATING UNL FOURIER FEATURES IN PRACTICE

In practice the pattern is given as a binary bitmap, without any a priori knowledge about the intrinsic mathematical definition of the pattern curves. The pattern is ap-

proximated by all linear curve segments

$$z(t) = (x_1 + t(x_2 - x_1)) + j(y_1 + t(y_2 - y_1)) \quad t \in (0, 1)$$

between two neighboring pixels of the pattern

$$z_1 = x_1 + jy_1 \text{ and } z_2 = x_2 + jy_2.$$

Hence once again an analytical description is available. In order to be able to define a line segment a thinning of the original pattern is necessary, e.g. by the algorithm in [27]. The centroid of the pattern is now the mean of all pattern coordinates: $O = (x_i, y_i)$, $i = 1 \dots \# \text{pixels}$.

The pattern is scanned once. If a pair of pixels is found the line segment is transformed by EQ3:

$$U(z(t)) = \frac{\| (x_1 + t(x_2 - x_1) - O_x) + j(y_1 + t(y_2 - y_1) - O_y) \|}{M} + j \times \text{atan} \left(\frac{y_1 + t(y_2 - y_1) - O_y}{x_1 + t(x_2 - x_1) - O_x} \right)$$

The parameter t is discretized into sufficiently small steps in the interval $(0,1)$. The only restriction is that no gaps in the transformed pattern may appear.

The DC component of the pixel matrix represents the basic signal energy. This is the Fourier magnitude for both polar parameters equal to zero: $F(0,0)$. This magnitude has always the maximum values of all magnitudes. Consequently we normalize all other magnitudes by dividing them by $F(0,0)$.

LIMITATIONS

The method assumes an ideal segmentation if the source images are non-binary, but so do other well established methods as well, e.g. 1-D Fourier descriptors. On the other hand patterns are often a priori only black and white, e.g. printed digits. In this case no segmentation is necessary. See [31] for an application of the UNL transform to handwritten characters.

The UNL Fourier Features are sensitive to occlusions. The set of parametric functions is changed abruptly if parts of the whole object are missing. Consequently the centroid and the UNL transform of the original pattern change completely and the UNL Fourier Features yield garbage values.

EXPERIMENTAL RESULTS

We present the results for a real world pattern classification task. A set of objects is presented to a standard 2-D vision system. The digital grey level images are segmented into binary images. From there the contours of the objects are extracted. The contour data is stored in files which are transferred to a workstation where higher level processing is performed.

Hardware: An industrial general purpose imaging system is used for frame grabbing and early vision. The Magiscan2 from Joyce Loebel [28] receives the analogous image signal from a conventional CCD-camera (Panason-

ic: Model WV-1500/B) and digitizes it to a 6 bit 512x512 pixel matrix. The vision system performs low level imaging to extract the 1 bit binary shape image. The Magiscan2 is controlled by a PC/AT. The binary contours are moved to a DECstation 3100 from Digital Equipment running under UNIX.

Contour Extraction: Images which failed to be segmented properly were purged, e.g. when holes appeared where the object has none or parts of the background were segmented into the object area. A mixture set of ideal 2-D paper sheet and other 3-D objects was used (fig. 3). The camera position was varied from an initial distance of 0.6m with 1cm increases. The segmentation was histogram based. After the segmentation, the horizontal chord segments were scaled by a constant linear factor of 1.088 in order to try to compensate for the distortion of a general purpose TV camera. The system calls of the vision system were used to extract the boundaries of the objects respectively of their holes. The whole set of images for one class was stored in a file and then transferred to the workstation.

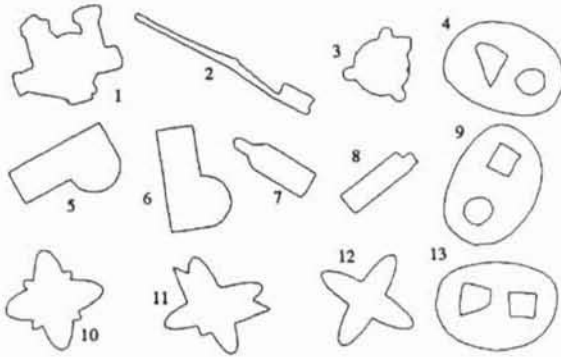


Fig. 3 The Benchmark Universe

Feature and Classifier Model: The original resolution of 512 was normally scaled down to a smaller resolution in the workstation environment. The preprocessing steps (thinning) were performed and finally the discrete UNL Transform of the pattern. The resolution of the polar coordinate system was the same as for the Cartesian coordinate system. Finally the UNL Fourier Features were calculated.

For each feature a Gaussian probability distribution was assumed. Hence the mean and standard deviation for each feature were estimated from the available training samples.

The totality of the samples was aleatorily split into 75% training data and 25% test data. The total number of samples per class were about 200. Over the whole feature pool a preselection was done to divide them into "good" and "bad". This quality label was based on an average interclass distance measure using the following heuristics:

$$Quality = \frac{2}{c(c-1)} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[1 - \frac{(\sigma_i + \sigma_j)^2}{(\mu_i - \mu_j)^2} \right] \quad EQ4$$

where c is the number of classes and μ_j and σ_j are the mean

and standard deviation for class i . The heuristics is the average interclass distance for all possible class pair combinations. The interclass distance between two classes is based on Chebychev's inequality. This formula results in a ranking between 0% and 100% for each feature. The heuristics makes the simplifying assumption that features are mutually independent which for a multivariate distribution is not true. On the other hand it allows to give a good rule of thumb for the potential of one feature to separate classes. Thus it is possible to order the features following this quality criteria.

For the preselected pool the values of all feature vectors were calculated. The feature selection based on multivariate analysis of covariance [29] finally determined which features were used for classification. The feature selection was initialized with that preselected feature that had the highest quality. Then the next feature was selected, joined to the selected pool and so on.

The number of selected features was based on a minimal estimated error rate using the leave-one-out method [30].

A nearest-neighbor classifier was applied for identification. Euclidean distance measure was used to establish an decision making function to determine to which class the unknown sample belonged to. No data reduction over the whole set of samples was performed. That means that an unknown sample compared to each of the known training samples.

Experiment: The general purpose classification performance of the UNL Fourier Features is tested. The parameters and results are presented in table 1. The samples were scaled from a resolution of 512 down to 256. The 13 classes were represented by 1928 samples. The pictures were taken from 10 different camera positions. The number of the preselected features was fixed at 100. From this pool 8 features were selected for the universe of objects under observation which were used for training and classification. The estimated recognition rate was 100%. For the 481 test samples the apparent recognition rate was also 100%.

It can be observed that the classifier performs well for the symmetric objects 5 and 6. Also the very similar objects 4, 9 and 13 are separated well (they differ only in one hole).

APPENDIX

A. Centroid of a finite set of parametric curves.

The common centroid O of an object (t) which is composed by n smooth parametric curves

$$z_i(t) = x_i(t) + jy_i(t), \quad i = 1 \dots n \quad \text{is:} \quad EQ-A1$$

$$O = (O_x, O_y) = \frac{\sum_{i=1}^n \int_0^1 z_i(t) \| \dot{z}_i(t) \| dt}{\sum_{i=1}^n \int_0^1 \| \dot{z}_i(t) \| dt}$$

The condition of smoothness is necessary to derive the

curve at every point. Furthermore it implies that the pattern is broken up into several curves at non continuous points.

B. Proof of Lemma 1.

Translation: Substitute $z_i(t)$ by $z_i(t) + \Delta t$ and calculate the new values for O, M , taking into account that Δt does not depend on t . Then EQ3 yields invariant values for $U(z_i(t))$.

Scaling: Multiply $z_i(t)$ by the positive real scalar a . Analogous case to translation.

Rotation: Substitute $z_i(t)$ by $\exp(j\phi)z_i(t)$ where ϕ is an angle by which the pattern is rotated about O . A rotation about any point can be split up into a translation to O , the rotation and inverse translation. Calculate the new values for O, M . It can be proved that for the polar angle θ

$\tan(\theta(z(t)) \exp(j\phi)) = \tan\theta(z(t) + \phi)$ which is equivalent to a cyclic shift of the transformed pattern.

C. Image of a curve.

$$I(z(t)) = \{ p \in C \mid z(t) = p, \forall t \in (0, 1) \}$$

This pointset is what we normally consider as the curve. It can be understood as an *instantiation* or *materilization* of the curve by variation of its parameter in the respective interval $(0,1)$.

REFERENCES

- [1] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Addison Wesley Publishing Corporation, 2.Edition, 1987.
- [2] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.
- [3] K. S. Fu, R. C. Gonzalez and C. S. G. Lee, *Robotics: Control, Sensing, Vision, and Intelligence*. McGraw-Hill, New York, 1987.
- [4] T. Pavlidis, "SURVEY-A Review of Algorithms for Shape Analysis," *Computer Graphics and Image Processing*, vol. 7, pp. 243-258, 1978.
- [5] ———, "Algorithms for Shape Analysis of contour and waveforms," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. PAMI-2, no. 4, pp. 301-312, July 1980.
- [6] D. H. Ballard and C. M. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.
- [7] H. Freeman, "Computer processing in line drawing images," *ACM Computer Surveys*, vol. 6, no. 1, pp. 57-98, March 1974.
- [8] J. Sklansky, R. L. Chazin and B. J. Hansen, "Minimum-Perimeter Polygons of Digitized Silhouettes," *IEEE Trans. Computers*, vol. C-21, no. 3, pp. 260-268, 1972.
- [9] T. Pavlidis, "Polygonal approximations by Newton's method," *IEEE Trans. Computers*, vol. C-26, no. 8, pp. 800-807, Aug. 1977.
- [10] C. T. Zahn and R. Z. Roskies, "Fourier Descriptors for plane closed curves," *IEEE Trans. Computers*, vol. C-21, pp. 269-281, 1972.
- [11] E. Persoon and King-Sun Fu, "Shape Discrimination using Fourier Descriptors," *IEEE Trans. Systems, Man, and Cybernetics*, vol. SMC-7, No. 3, pp. 170-179, 1977.
- [12] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Trans. Information Theory*, vol. IT-8, pp. 179-187, Feb. 1962.
- [13] A. P. Reeves, R. J. Prokop, S. E. Andrews and F. P. Kuhl, "Three-Dimensional Shape Analysis Using Moments and Fourier Descriptors," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 10, no. 6, pp. 937-943, Nov. 1988.
- [14] T. H. Reiss, "The Revised Fundamental Theorem of Moment Invariants," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 13, no. 8, pp. 830-834, Aug. 1991.
- [15] C.-H. Teh and R. T. Chin, "On Image Analysis by the Methods of Moments," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 10, no. 4, pp. 496-513, July 1988.
- [16] A. Khotanzad and Y. H. Hong, "Invariant Image Recognition by Zernike Moments," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 12, no. 5, pp. 489-497, May 1990.
- [17] E. L. Schwartz, "Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception," *Biol. Cybern.*, vol. 25, pp. 181-194, 1977.
- [18] ———, "A quantitative model of the functional architecture of human striate cortex with applications to visual illusion and cortical texture analysis," *Biol. Cybern.*, vol. 37, pp. 63-76, 1980.
- [19] ———, "Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding," *Vision Res.*, vol. 20, pp. 645-669, 1980.
- [20] L. Massone, G. Sandini and V. Tagliasco, "'Form-Invariant' Topological Mapping Strategy for 2D Shape Recognition," *Computer Vision, Graphics and Image Proc.*, vol. 30, pp. 169-188, 1985.
- [21] H. Wechsler and G. L. Zimmerman, "2-D Invariant Object Recognition Using Distributed Associative Memory," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. PAMI-10, no. 6, pp. 811-821, Nov. 1988.
- [22] W. L. Reber and J. Lyman, "An Artificial Neural System Design for Rotation and Scale Invariant Pattern Recognition," in *Proc. IEEE First Int. Conf. on Neural Networks*, San Diego, CA, 1987, vol. IV, pp. 277-283.
- [23] K. S. Fu, *Syntactic Pattern Recognition and Application*. New York: Academic, 1972.
- [24] R. L. Kashyap and R. Chellappa, "Stochastic Models for closed boundary analysis: Representation and reconstruction," *IEEE Trans. Information Theory*, vol. IT-27, pp. 627-637, Sept. 1981.
- [25] Y. He and A. Kundu, "2-D Shape Classification Using Hidden Markov Model," *IEEE Trans. Pattern Analysis and Machine Intell.*, vol. 13, no. 11, pp. 1172-1184, Nov. 1991.
- [26] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer*. 1988.
- [27] T. Y. Zhang and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns", in *Comm. ACM*, vol. 27, NO. 3, pp. 236-239, 1984.
- [28] Joyce Loeb Inc. Manual of the *Magiscan2 Vision System*, 1986.
- [29] W. W. Cooley and P. R. Lohnes, *Multivariate Data Analysis*. New York: Wiley, 1971.
- [30] P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*. Prentice/Hall Int., London 1982.
- [31] T. W. Rauber and A. S. Steiger-Garçon, "Shape Description by UNL Fourier Features—An Application to Handwritten Character Recognition," in *Proc. 11th IAPR Int. Conf. on Pattern Recognition*, The Hague, The Netherlands, August 1992, vol. II, pp. 466-469.

