Character Rotation Absorption Using a Dynamic Neural Network Topology: Comparison With Invariant Features

Christophe Choisy[1], Hubert Cecotti[1], and Abdel Belaïd[1]

LORIA
Campus Scientifique, BP 239
54506 Vandoeuvre-ls-Nancy cedex, France

Abstract. This paper treats on rotation absorption in neural networks for multi-oriented character recognition. Classical approaches are based on several rotation invariant features. Here, we propose to use a dynamic neural network topology to absorb the rotation phenomenon. The basic idea is to preserve as most as possible the graphical information, that contains all the information. The proposal is to dynamically modify the neural network architecture, in order to take into account the rotation variation of the analysed pattern. We use too a specific topology that carry out a polar transformation inside the network. The interest of such a transformation is to transform the rotation problem from a $2D$ problem to a $1D$ problem, that is easier to treat. These proposals are applied on a synthetic and on a real EDF[1] base of multi-oriented characters. A comparison is made with Fourier and Fourier-Mellin invariants.

1 Introduction

There are two main opposed approaches in character recognition: the first one tries to find a set of representative features for each class [6], as the second one tries to learn all the possible character deformations and variations [9]. If the first method could be efficient with a low cost for relatively clean data, the second one has proved its superiority on variable characters like handwritten digits [8].

The backdraw of this method is that it requires to dispose of all the possible pattern deformations in the training set. When dealing with relatively straight characters it is possible to create such a set [9]. But for multi-oriented characters the problem is multiplied by all the rotated position possibilities, leading to a very huge database. Moreover, the number of model parameters will also greatly increase, leading to a very complex system. This is why usually one proposes to extract some invariant features to represent each class [1, 10, 4]. But as said before, this method is less efficient than learn all the possibilities, because some information is lost in the features extraction.

In order to not loose any information, we avoid the intermediate feature extraction step, working directly on the image. Furthermore, as it is impossible to learn all the

[1] EDF: the French national company of electricity

Choisy C., Cecotti H. and Belaïd A. (2004).
Character Rotation Absorption Using a Dynamic Neural Network Topology: Comparison With Invariant Features.
In *Proceedings of the 4th International Workshop on Pattern Recognition in Information Systems*, pages 90-97
DOI: 10.5220/0002683500900097

multi-oriented variations, we propose to absorb the rotation by adapting the system topology to the deformation produced by the rotation. The interest of such a technique is to obtain model and training set conditions similar to those of straight characters approaches. Thus the model complexity and the database size remain reasonable.

Neural networks offer the possibility to dynamically modify their topology making easy the implementation of such approach. The purpose of this paper is to show how to dynamically modify the network topology by acting on the network links without disturbing the training efficiency. The principle of the dynamic changing is to act in a same manner on all the links of all the neurons on a specific layer. This guarantees to consistently distribute the error during the learning step.

In order to facilitate the dynamic network adaptation, it is usefull to proceed to a polar transformation. This makes it possible to pass from a $2D$ problem to a $1D$ problem. In order to disturb as less as possible the network process, we propose to code the polar transformation as a specific neural topology. This leads the network deciding itself about the interest of each pixel in the transformation.

2 System overview

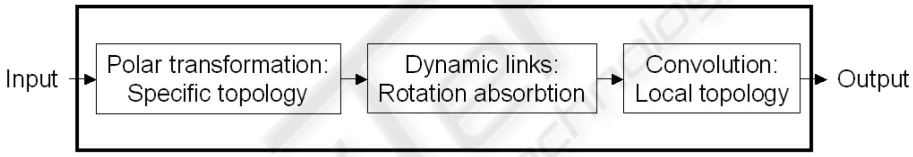
Input
Polar transformation: Specific topology
Dynamic links: Rotation absorbtion
Convolution: Local topology
Output

Fig. 1. System overview

Our system is based on a multi-layer perceptron with several specificities. It is designed to be as most as possible preprocessing-free. We only scale the images according to the neural network input size. Figure 1 shows the different steps in our neural network.

The input layer takes the gray-level pixels of the pattern. The link between the first and the second layer are organized in order to complete a polar transformation of the image: thus the second layer can be assimilated to the polar version of the input image. During learning step, the link weights are re-estimated in order to take into account pixel interest in the transformation. This polar transformation allows to transform the problem of image rotation into a problem of shift along the angle axis.

The second and third layers are connected with dynamic links. Modifying these links allows to adapt to the shift. Thus the information will be transferred to the next layers as if the input image was straight.

A layer with local convolutions is added to refine the image analysis. The last layer gives the output probabilities for the different classes.

3 The polar transformation

The problem of rotation absorption leads often to transform the image into polar coordinates in a first step [7, 3]. Indeed, it is easier to deal with the polar form, because the rotation problem is transformed in a translation problem: thus, a $2D$ deformation becomes a $1D$ deformation.

Consider an image I_c of size $X * Y$ defined in Cartesian coordinate $I_c(x; y)$ where $0 \leq x \leq X-1$ and $0 \leq y \leq Y-1$. If the image I_c is rotated through angle $\alpha \in [0; 2\pi]$, the new image is defined as $I'_c(x'; y') = I_c(x * cos(\alpha); y * sin(\alpha))$.

Let G be the gravity center of the image. Consider the image in polar coordinate $I_p = (r, \theta)$: the relationship between I_p and the rotated image though angle α is $I'_p(r'; \theta') = I_p(r; \theta + \alpha)$. r is defined as the distance between G and the point. $\theta \in [0; 2\pi]$ corresponds to the angle. Thus the angle variations can be treated as a shift along the θ axe. Several ways are proposed to perform this transformation.

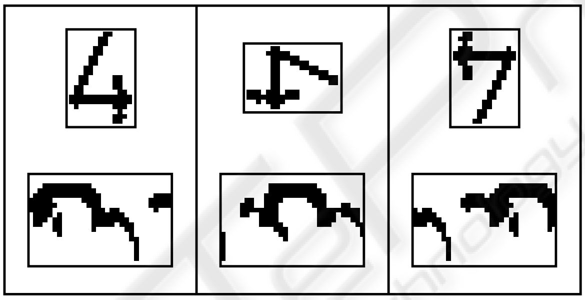

Fig. 2. Rotation-to-Shift effect of the polar transformation: Top: original image, Bottom: transformed image

3.1 Goshtasby transformation improvements

Let R be the maximum radius of the pattern. R is defined as the maximum Euclidean distance between the center of mass and the furthest gray pixel (not isolated, to avoid taking noise as pixel reference). The Goshtasby method [7] is to draw n concentrically circles around G with radius $R * i/n$; $i \in \{1, .., n\}$. Each circles is cut out m parts. In this proposal only the pixels in the intersection of circles and the partition of angles are chosen, to build the $n * m$ polar shape matrix.

With this kind of description, the original image is not fully represented in the transformed one; thus we first propose to use the mean of the sector pixels rather than only the intersecting points.

However surfaces are growing with the radius, thus the parts near the center contain few information as the furthest keep a lot of information. Thus we propose another approach, based on equal sector surfaces.

Let r be the portion's step of the radius, πr^2 is the surface of the inner ring. Consider the ring n, the surface S_n of this ring is defined as $\pi((n+1)r))^2 - \pi(nr)^2$, $S_n = (2n+1) * \pi r^2$. By dividing the n^{ht} ring into $2n+1$ parts, sector surfaces are equal for each ring on the image. The number of sectors of the polar image for n rings is $\sum_{i=0}^{i=n-1}(2i+1)$.

This approach allows a better repartition of the pixels in the transformed image, as shows figure 3. Some tests on multi-oriented character recognition show about 5% improvement for the equal surfaces approach.

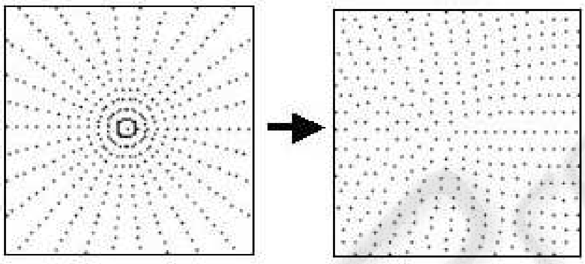

Fig. 3. Sector center repartition considering classical (left) and equal surface (right) transformation

3.2 Integration of the polar transformation in a Neural Network

It is classical to give a topology to a neural network, in order to search elementary patterns and decrease the network complexity. The advantage is to let the network learn itself what is the contribution of each pixel in these patterns. In the same idea, we propose to integrate the polar transformation in the neural network through a specific topology.

The transformation is made through the links between two layers. The first one represents the Cartesian image whereas the second one represents the polar image. The principle is similar to the mathematical transform, except than pixels are not directly moved: the links are fixed in such a manner than the input neuron position is taken in Cartesian coordinates, and the destination neuron position correspond to the polar destination of the pixel. The figure 4 illustrate the two layers with some links, for the classical polar transformation.

The classical and equal surfaces transformation were tested using appropriated specific topologies. But for the following work we only use the classical transformation, due to an interesting property: each column of the transformed image corresponds to an angle portion (see figure 4). For the equal surfaces transformation the shift problem will vary according to the radius (see figure 5): this leads to a more complex dynamic procees, needing some interpolation depending of the observed ring.

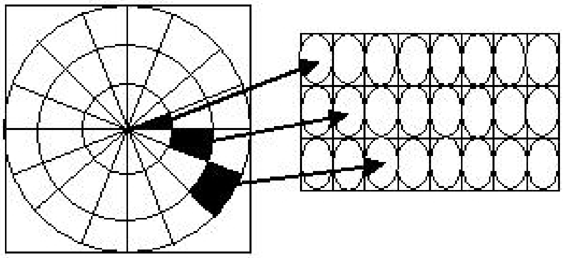

Fig. 4. Left: Layer 0, Input Image ; Right: Layer 1, image in polar coordinate. Rows represent set of links between neurons

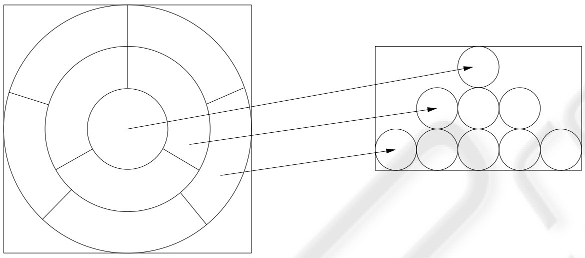

Fig. 5. Equal surface analysis needs a pyramidal layer struct, that leads to a more complex dynamic process.

4 Dynamic topology for the process

The use of a dynamic topology is a extension of the convolution idea for neural network. As said in the introduction, an optimal learning needs to have seen all the possible pattern distortions. For multi-oriented data, that conduces to multiply the data number by all the possible angles. This leads to a very big database, and needs a big neural network to deal with so many parameters. It could be possible to achieve such a system, but it is not realistic to use it.

After the polar transformation, the neural network has to deal with a problem of shift along the angle axe. Our proposal is to correct the shift error, in order to bring the image back to its straight position. Thus it is not needed to learn all the possible rotation possibilities, but only the straight samples. This allows to have a network complexity similar to the classical other applications.

The principle of the dynamic links is to change the input neuron of each link according to the shift error. The intuitive idea is shown in figure 6. The objective is that the first link of each neuron will always be connected to the real first column of the polar image. If the image is shifted, the link is dynamically moved to the position of the first column. This movement is synchronized for all the dynamic links, according to the observed sector. On an output neuron line, all the neurons observe the same input line, and the links are moved synchronously. It allows to have classically several output neurons for the same input neuron group.

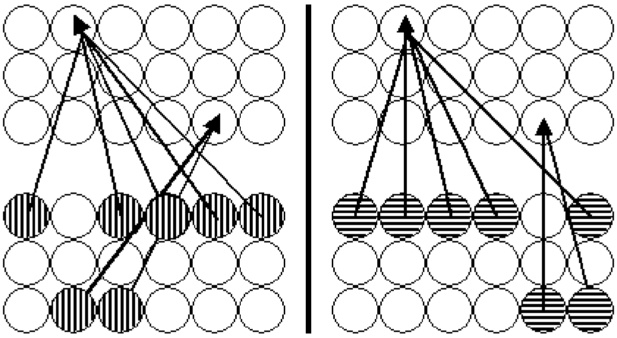

Fig. 6. Dynamic topology. Left: Input image straight, Right: Input image oriented

This introduce the notion of order in the links: if the training and testing calculation give the same role to each link, the dynamic part differentiate them according to the location they analyze on the image.

5 Implementation and experimental results

Our system has been tested on two databases. The first one is synthetic: its interest is to validate the proposed approach and to clearly compare it to invariant feature extraction. The second database is a real database of multi-oriented and multi-sized characters, extracted from EDF[2] technical maps.

The synthetic database was builded from different orientations of characters from one font. It has 62 classes. The training base considers 12 orientations by 30 degrees step beginning in 0. The testing base aslo considers this step, but beginning in 15 degrees. Considering all classes, it leads to 744 patterns to train, and same to test.

The real database was restricted to 30 major classes, as proposed in another work [1]. It contains 15268 samples, divided in 80% for training and 20% for testing. The distribution of the characters are not homogeneous. An orientation approximation made by EDF while character strings are extracted from maps, shows that half of the character have an estimated slope lower than 5^o. It confirms one of our hypothesis: a database with enough samples for all the possible orientations will be very huge.

In our first tests, we used the orientation information to dynamically adapt the topology. The objective is to validate the interest of such an approach. Obviously, the perspective of this work is clearly to automatically decide this dynamic modification. Results are given considering straight rectified characters, dynamic topology, Fourier-Mellin invariants, Fourier invariants, and multi-oriented characters. Fourier invariants are extracted fater polar transformation; assuming a $16x32$ polar image, it gives 512 invariants. Fourier-Mellin invariant number depends on the precision of the analysis: the parameter number is indicated.

Synthesis database results are given in table 1.

[2] EDF: the French national company of electricity

Table 1. Synthesis database: recognition rate for 62 classes

Straight rectified characters	92.47%
Dynamic Topology	88.44%
Fourier-Mellin (33 features)	83.20%
Fourier features	81.89%
Multi-oriented characters	45.83%

Table 2 Fourier and Fourier-Mellin results when samples are grouped into rotation similar classes.

Table 2. Synthesis database with 44 grouped classes

Fourier-Mellin (33 features)	86.83%
Fourier features	87.90%

Results show the superiority of the dynamic topology approach. The difference between this approach and straight rectified character results, resides in the fact that rectification is made precisely at the rotation degree, when dynamic topology absorbs only with a variation of +/-5 degrees. The small superiority of Fourier when considering grouped classes can be explained by its higher parameter number regarding to Fourier-Mellin approach.

Results on the EDF real database are given in table 3. The score for multi-oriented characters is better than for invariant features. This phenomenon can be explained by the fact that character slopes are badly distributed, and should conduce to orientation clusters. The superiority of the dynamic topology can be explained by the approximative orientation given by EDF: the variation induced by the dynamic topology implementation allows to better distribute the samples in the topology variation range.

Table 3. Real database: recognition rate for 30 classes

Straight rectified characters	87.31%
Dynamic Topology	88.97%
Fourier-Mellin (33 features)	68.85%
Fourier-Mellin (119 features)	76.72%
Fourier features	73.16%
Multi-oriented characters	76.89%

All these results prove that invariant features cannot compete with graphic information. This shows the interest of a dynamic topology. Let us precise that the 30 classes selected for this test seem not to be the best choice to really reduce the rotation confusion problem. We are trying another clusters in order to have more differenciated Fourier-Mellin features.

6 Conclusion

We have proposed a new methodology to absorb image rotations in a neural network without extracting invariant characteristics. This method is based on a dynamic neural network topology, and gives encouraging results when applied on multi-oriented and multi-scaled images. Compared to rotation invariant feature extraction, this approach proved its interest, showing that graphical information can perform better than features extraction.

Our aim is to propose in a future work a decisional process, that will be able to decide of the dynamic link modification. Another work concerns the polar transformation: the best version induces a more complicated dynamic topology process, that we have to implement; some tests show that this new version could improve recognition from about 5%.

References

1. Adam, S., Ogier, J.M., Cariou, C. Mullot, R., Gardes, J., Lecourtier, Y.: Fourier-mellin based invariants for the recognition of multi-oriented and multi-scaled shapes : application to engineering drawings analysis. Machine perception and artificial intelligence (2000) 42:132–147.
2. Brandt, R.D., Lin, F.: Representations that uniquely characterize images modulo translation, rotation, and snewblockcaling. Pattern Recognition Letters (1996), 9(17):1001–1015.
3. Bui, T., Chen, G., Roy, Y.: Translation-invariant multiwavelets for image de-noising. Proc. 3rd World Multi-Conference on Circuits, Systems, Communications and Computers (1999).
4. Derrode, S., Ghorbel, F.: Robust and efficient fourier-mellin transform approximations for gray-level image reconstruction and complete invariant description. Proc. 3rd World Multi-Conference on Circuits, Systems, Communications and Computers.(2001), 83(1).
5. Deseilligny, M., Le Men, H., Stamon, G.: Characters string recognition on maps, a method for high level reconstruction. ICDAR (1995), 249–252.
6. Dur Trier, O. Jain, A.K., Taxt, T.: Feature extraction methods for character recognition – a survey. Pattern Recognition (1996), 4(29):641–662.
7. Goshtasby, A.: Description and discrimination of planar shapes using shapes matrices. IEEE Transactions of Pattern Recognition and Machine Intelligence (1985), PAMI-7(6):738–743.
8. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognitionntion. Intelligent Signal Processing, 306–351, 2001.
9. Simard, P., Steinkraus, D., Platt, J.: Best practice for convolutional neural networks applied to visual document analysis international conference on document analysis and recogntion. ICDAR, IEEE Computer Society (2003), 958–962.
10. Zhenjiang, M.: Zernike moment-based image shape analysis and its application. Pattern Recognition Letters (2000), 2(21):169–177.