

A Shape Descriptor Based on Trainable COSFIRE Filters for the Recognition of Handwritten Digits

George Azzopardi and Nicolai Petkov

Johann Bernoulli Institute for Mathematics and Computer Science
University of Groningen, The Netherlands
{g.azzopardi,n.petkov}@rug.nl

Abstract. The recognition of handwritten digits is an application which has been used as a benchmark for comparing shape recognition methods. We train COSFIRE filters to be selective for different parts of handwritten digits. In analogy with the neurophysiological concept of population coding we use the responses of multiple COSFIRE filters as a shape descriptor of a handwritten digit. We demonstrate the effectiveness of the proposed approach on two data sets of handwritten digits: Western Arabic (MNIST) and Farsi for which we achieve high recognition rates of 99.52% and 99.33%, respectively. COSFIRE filters are conceptually simple, easy to implement and they are versatile trainable feature detectors. The shape descriptor that we propose is highly effective to the automatic recognition of handwritten digits.

Keywords: COSFIRE, feature, descriptor, trainable, recognition, shape.

1 Introduction

Handwritten digit recognition is a challenging task with various applications, such as bank cheque processing [23] and postal mail sorting [2]. Feature extraction plays an important role in the effectiveness of such systems. A number of methods have been proposed, including Zernike moments [3], direct matching [9], Fourier descriptors [12], geometric moment invariants [21], shape context [4], zoning [8] and biologically motivated features [10]. For a thorough overview of recent advances in the recognition of Western Arabic and Farsi handwritten digits we refer the reader to [16,22] and [20,5], respectively.

We use trainable COSFIRE feature detectors of the type introduced in [1]. They are automatically configured to be selective for parts of given training digit images. The automatic configuration process extracts information about the local geometric arrangement of contour segments. COSFIRE filters are inspired by the properties of a specific type of shape-selective neuron in area V4 of visual cortex, which exhibit selectivity for parts of (curved) contours or for combinations of line segments [18].

The response of a COSFIRE filter in a given point is computed as a function of the shifted responses of simpler (in this case orientation-selective) filters. Using shifted responses of simpler filters, such as Gabor filters that we use in this study, corresponds to combining their respective supports at different locations to obtain a more sophisticated filter with a bigger support. The specific function that we use here to combine the responses of Gabor filters is geometric mean, essentially multiplication, which has specific advantages regarding shape recognition and robustness to contrast variations. One consequence is that a COSFIRE filter produces a non-zero response only when all constituent parts of a pattern of interest are present.

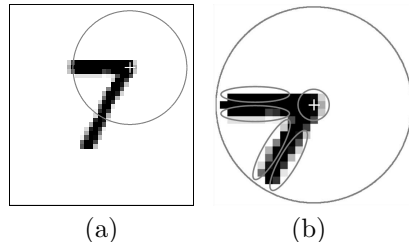
The rest of the paper is organized as follows: in Section 2 we present the COSFIRE filter and demonstrate how it can be configured and used to detect features of handwritten digits. We also show how the responses of multiple COSFIRE filters can be used to form a shape descriptor of a handwritten digit. In Section 3, we evaluate the effectiveness of our method on two data sets of handwritten digits: Western Arabic (MNIST) [15] and Farsi [13]. We provide a discussion in Section 4 and draw conclusions in Section 5.

2 Proposed Method

2.1 Overview

Fig. 1a illustrates a handwritten digit with an encircled feature. We use this local pattern (shown enlarged in Fig.1b) to automatically configure a COSFIRE filter that will respond (with certain tolerance) to the same and similar patterns.

Fig. 1. (a) The circle indicates a feature that is selected from a given training digit image. (b) Enlargement of the selected feature. The ellipses represent the dominant orientations around the specified point of interest that is marked by a ‘+’ marker.



The four ellipses shown in Fig.1b illustrate the dominant orientations of edges in the surroundings of the point of interest. We use 2D Gabor filters to detect these orientations. The small central circle represents the overlapping supports of a group of such Gabor filters. The response of a COSFIRE filter is computed by combining the responses of the concerned Gabor filters by geometric mean.

The preferred orientations of these Gabor filters and the locations at which we take their responses are determined by a configuration process which automatically analyses the specified pattern. Taking the responses of Gabor filters at different locations around a point can be implemented by shifting the responses of these Gabor filters by different vectors before using them for the pixel-wise evaluation of a (geometric mean) function that gives the COSFIRE filter output.

In the following, we explain the automatic configuration process of a COSFIRE filter that determines which responses of which Gabor filters in which locations need to be multiplied in order to obtain the output of the filter. Subsequently, we introduce a feature vector as a shape descriptor of a handwritten digit that is formed by the responses of multiple COSFIRE filters. Finally, we apply support vector machines (SVM) to such feature vectors to achieve handwritten digit recognition.

2.2 Detection of Dominant Orientations by 2D Gabor Filters

We use 2D antisymmetric Gabor filters to detect the edges of the pen strokes in the handwritten digit images. We denote by $g_{\lambda,\theta}(x,y)$ the half-wave rectified response of a Gabor filter of preferred wavelength λ and orientation θ to a given input image. Such a filter has a number of other parameters, including spatial aspect ratio, bandwidth and phase offset, which we set as proposed in [19].

We apply a bank of Gabor filters with one wavelength $\lambda = 2\sqrt{2}$ and 16 orientations ($\theta \in \{0, \frac{\pi}{16}, \dots, \frac{15\pi}{8}\}$) on handwritten digit images of width 28 pixels.

We threshold the responses of Gabor filters at a given fraction $t_1 = 0.1$ of the maximum response of $g_{\lambda,\theta}(x,y)$ across all combinations of values (λ, θ) used and all positions (x,y) in the image. The threshold value $t_1 = 0.1$ is sufficient to suppress the undesirable responses of Gabor filters to background noise.

2.3 Detection of Digit Parts by COSFIRE Filters

We use COSFIRE filters of the type proposed in [1] to detect digit parts. The automatic configuration process of a COSFIRE filter results in a set of four-tuples: $S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i = 1 \dots n_f\}$ where n_f stands for the number of dominant contour parts in a given pattern f . Each tuple of four parameter values $(\lambda_i, \theta_i, \rho_i, \phi_i)$ characterizes the properties of a given orientation of a contour part that is present in the specified pattern of interest: the parameters (λ_i, θ_i) represent a Gabor filter that is selective for the concerned contour part and (ρ_i, ϕ_i) are the polar coordinates of the location where that contour part lies with respect to the center of the COSFIRE filter.

We compute the response of a COSFIRE filter as follows. For each tuple with index i we first apply the Gabor filter with parameters (λ_i, θ_i) . Then, we blur the Gabor responses with a Gaussian function in order to allow for some tolerance in the position of the involved contour part. The standard deviation σ_i of such a blurring function grows linearly with the parameter ρ_i :

$$\sigma_i = \sigma_0 + \alpha\rho_i \tag{1}$$

where σ_0 and α are constant values. Finally, we shift the blurred Gabor responses with ρ_i pixels in the direction opposite to ϕ_i such that the responses of different Gabor filters meet at a point which we consider the center of the concerned COSFIRE filter. The output of a COSFIRE filter $r_{S_f}(x,y)$ is computed as the geometric mean of the blurred and shifted responses of the involved Gabor filters.

Rotation invariance is achieved by manipulating the set of parameter values in S_f , rather than by computing them from the responses to rotated versions of the original pattern. Using the set S_f that defines the concerned filter, we first

form a new set $\mathfrak{R}_\psi(S_f) = \{(\lambda_i, \theta_i + \psi, \rho_i, \phi_i + \psi) \mid \forall (\lambda_i, \theta_i, \rho_i, \phi_i) \in S_f\}$. Then, we define the rotation-invariant response as $\hat{r}_{S_f}(x, y) = \max_{\psi \in \Psi} (r_{\mathfrak{R}_\psi(S_f)}(x, y))$ for a given set Ψ of rotation increments.

2.4 Formation of a Shape Descriptor

We use the collective responses of multiple COSFIRE filters as a shape descriptor of a given handwritten digit. Such a description is inspired by the neurophysiological concept of population coding [18] where a given stimulus can be described by the pattern of neuronal activity that it invokes in a given cortical area. One can speculate that the response of a COSFIRE filter in a given point corresponds to the activity of a shape-selective V4 neuron. The responses of many different COSFIRE filters, possibly at different positions, to an input digit image would represent the collective activity of a group of shape-selective V4 neurons.

For a given training set of digit images, we first configure a number of COSFIRE filters for different parts of the handwritten digits. In the configuration step, we choose a random subset of training digit images from each digit class. We configure one COSFIRE filter for every selected digit with a local pattern that is randomly selected in the input image. For a randomly selected pattern to be valid and used for configuration it should result in the configuration of a filter that combines at least four Gabor filter responses, otherwise we choose a different location. Such a restriction is required in order to avoid the selection of small patterns which may consequently result in filters with low discriminative power. The configuration of the COSFIRE filters is determined by four values of ρ ($\rho \in \{0, 3, 7, 12\}$).

Every training digit is then described by a feature vector where each element corresponds to the maximum response of a COSFIRE filter across all locations in the input image. For instance, with k filters, a digit is described by a vector of k elements. A value of zero means that the corresponding filter is not activated for the presented digit image, while a non-zero response corresponds to the activation strength of the filter. We then use the resulting feature vectors of the training set to train an all-pairs multi-class (with majority vote) SVM classifier¹ with a linear kernel.

3 Experimental Results

In the following, we evaluate the effectiveness of the proposed approach on two data sets of handwritten digits: MNIST data set of Western Arabic digits² [15] that comprises 60000 training digits and 10000 test digits, and a data set of Farsi digits³ [13] which includes 60000 training digits and 20000 test digits. The digits in the MNIST data set are given as grayscale images of size 28×28 pixels. The Farsi digits are given as binary images of varying sizes. We use bicubic interpolation to resize the images of the Farsi digits to a width of 28 pixels.

For each data set, we perform a number of experiments with different values of the threshold parameter t_1 ($t_1 \in \{0.05, 0.1, 0.15\}$). The values of the other two

¹ We use the SVM^{light} implementation: <http://svmlight.joachims.org/>

² The MNIST data set can be downloaded from <http://yann.lecun.com/exdb/mnist>

³ The Farsi data set can be downloaded from <http://www.modares.ac.ir/eng/kabir>

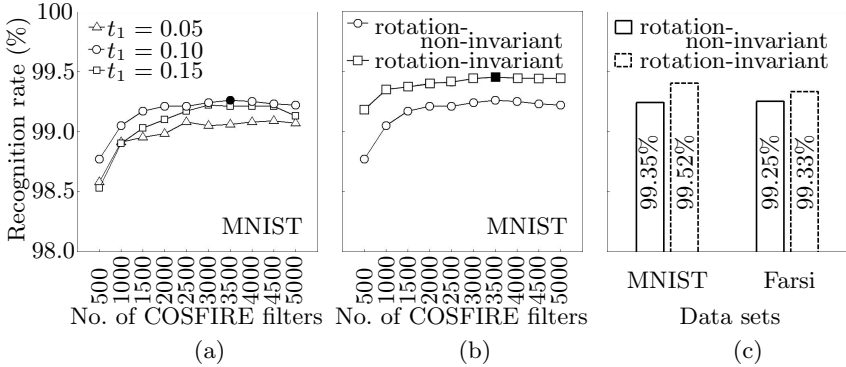


Fig. 2. Experimental results. (a) The three plots show the recognition rates for the MNIST data set achieved for different values of the threshold t_1 as a function of the number of COSFIRE filters used. Here, we use $\sigma_0 = 0.05$ and $\alpha = 0.80$. The filled-in circle represents the maximum recognition rate of 99.26% achieved for $t_1 = 0.1$ with 3500 filters. In these experiments the COSFIRE filters are used in rotation-non-invariant mode. (b) Performance comparison between a set of COSFIRE filters and $t_1 = 0.1$ that are first applied in rotation-non-invariant mode and then in a partially rotation-invariant mode with the best parameters $\sigma_0 = 0.07$ and $\alpha = 0.85$. Here, partial rotation-invariance is based on five values of the rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The performance improves with partial rotation-invariant filters that achieve a recognition rate of 99.52% (shown as a filled-in square). (c) The best recognition rates achieved for the MNIST and Farsi data sets with 3500 rotation-non-invariant and partially rotation-invariant COSFIRE filters with parameters $\sigma_0 = 0.07$ and $\alpha = 0.85$.

parameters are kept fixed for all experiments: $\sigma_0 = 0.05$ and $\alpha = 0.80$. For every value of t_1 we run an experiment by configuring up to 5000 COSFIRE filters. We repeat such an experiment five times and report the average recognition rate. Repetition of experiments is required in order to compensate for the random selection of digits and the random selection of locations that are used to configure the concerned filters. In Fig. 2a we plot the recognition rates that we achieve for the MNIST data set for different values of the threshold t_1 and for different number of COSFIRE filters used. We achieve a maximum recognition rate of 99.26% with 3500 COSFIRE filters for $t_1 = 0.1$, $\sigma_0 = 0.05$ and $\alpha = 0.80$.

Then, we run another set of experiments with different values of the parameters σ_0 ($\sigma_0 \in \{0.01, 0.03, \dots, 0.15\}$) and α ($\alpha \in \{0.70, 0.75, \dots, 1\}$). For the new set of experiments we use the value of the threshold parameter $t_1 = 0.1$, the one which performed best in the previous experiments. We achieve a maximum recognition rate of 99.35% for $\sigma_0 = 0.07$ and $\alpha = 0.85$ with 3500 COSFIRE filters.

For this set of experiments we apply the COSFIRE filters in rotation-non-invariant mode. One can observe, however, that some of the digits given in the two data sets differ slightly in orientation. We consider this fact and repeat the five experiments for the threshold $t_1 = 0.1$, $\sigma_0 = 0.07$ and $\alpha = 0.85$ (that contributed to the best performance so far), but this time applying the same COSFIRE filters in a partially rotation-invariant mode with five values of the

rotation tolerance angle ψ ($\psi \in \{-\frac{\pi}{4}, -\frac{\pi}{8}, 0, \frac{\pi}{8}, \frac{\pi}{4}\}$). The plots in Fig. 2b show that the performance that is achieved with the partially rotation-invariant filters is improved to a maximum recognition rate of 99.52% with 3500 filters. This means that the error rate is decreased by 26.15%.

We run similar experiments on the Farsi data set and achieve the best average recognition rate of 99.25% for $t_1=0.1$, $\sigma_0 = 0.07$ and $\alpha = 0.85$ with rotation-non-invariant COSFIRE filters and an average recognition rate of 99.33% with the same filters applied in a partially rotation-invariant mode. Fig. 2c illustrates these results next to the best results that we obtain for the MNIST data set.

4 Discussion

The high recognition rates that we achieve demonstrate the effectiveness of the proposed COSFIRE filters for the application at hand. The recognition rate of 99.52% that we achieve for the MNIST data set is comparable to the best existing approaches⁴. In particular, our method outperforms the shape context approach (99.37% in [4]), and three other approaches (94.20% in [17], 97.62% in [5] and 98.73% in [10]) that use biologically inspired feature detectors combined with a multi-layer perceptron (MLP) [17] and a linear SVM classifier [5,10]. The highest recognition rate achieved to date is 99.73% [7]. The approach used to achieve such a result extends the original training data set by elastically distorting the training images. For the Farsi data set, the recognition rate of 99.33% that we achieve is better than the best results ever reported in the literature. The COSFIRE filters outperform other biologically inspired feature detectors (96% in [5] and 99.1% in [10]) combined with a linear SVM classifier and also outperform the modified gradient technique which resulted in a recognition rate of 98.8% when combined with a multiple classifier system based on MLP classifiers [13].

Notable is the fact that we achieve the above results without performing any pre- and/or post-processing operations, neither do we use an extended training data set with elastic distortion. We apply the same methodology for both data sets, even though the corresponding numerals have different characteristics. This versatility is attributable to the fact that the COSFIRE filters are trainable, in that they can be automatically configured to any given patterns. This is in contrast to the best methods applied on the MNIST data set in which elaborate application-specific techniques have resulted from long-lasting research effort.

The proposed method contains an interesting aspect from a machine learning point of view. In traditional machine learning, the features to be used are fixed in advance and the machine learning aspect concerns the classification of observed feature vectors. If traditional machine learning is concerned with features at all, this is typically limited to the selection of predefined features or using them to derive ‘new’ features as (linear) combinations of the original ones. Examples are principle component analysis and generalized matrix learning vector quantization [6]. Traditional machine learning is typically not concerned with the question of how the features are defined. This aspect of the problem is, however,

⁴ A list of results obtained by state-of-the-art approaches is maintained at <http://yann.lecun.com/exdb/mnist/>

crucial for the success: almost any machine learning method will perform well with good features. The interesting aspect we would like to point out is that in the proposed approach the appropriate features are learned in the filter configuration process when a digit part is presented. We configure COSFIRE filters to be selective for local contour patterns that are randomly selected (with few restrictions) from given training images. In our experiments, we do not analyse the discriminative ability of the configured filters. This might cause the generation of filters that are selective to correlated patterns or to patterns with low distinctiveness. A feature selection method can therefore be incorporated in a machine learning algorithm, such as relevance learning vector quantization [11] or a support feature machine [14], in order to identify the most relevant filters.

It is worth noting that we do not use information about the mutual arrangement of the features obtained with different COSFIRE filters - a high value of the response of a given COSFIRE filter signals the presence of a given feature, but does not specify where this feature is in the image.

5 Conclusion

We demonstrated that the versatile COSFIRE filters are very effective feature detectors of handwritten digits. We achieved recognition rates of 99.52% and 99.33% for the MNIST and Farsi data sets, respectively. These results outperform other techniques that are based on biologically inspired features.

The proposed approach of combining the responses of a collection of COSFIRE filters to form a shape descriptor can be thought of as a general framework for many pattern recognition and machine vision problems.

References

1. Azzopardi, G., Petkov, N.: Trainable COSFIRE Filters for Keypoint Detection and Pattern Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(2), 490–503 (2013)
2. Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., Basu, D.K.: Recognition of Numeric Postal Codes from Multi-script Postal Address Blocks. In: Chaudhury, S., Mitra, S., Murthy, C.A., Sastry, P.S., Pal, S.K. (eds.) *PRMI 2009*. LNCS, vol. 5909, pp. 381–386. Springer, Heidelberg (2009)
3. Belkasim, S.O., Shridhar, M., Ahmadi, M.: Shape recognition using Zernike moment invariants. In: Chen, R.R. (ed.) *Conference Record. Twenty-Third Asilomar Conference on Signals, Systems and Computers (IEEE Cat. No.89-CH2836-5)*, vol. 1, pp. 167–171. IEEE, Naval Postgraduate Sch., San Jose State Univ. Conference Record. *Twenty-Third Asilomar Conference on Signals, Systems and Computers (IEEE Cat. No.89-CH2836-5)*, Pacific Grove, CA, USA, October 30–November 1 (1989)
4. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4), 509–522 (2002)
5. Borji, A., Hamidi, M., Mahmoudi, F.: Robust handwritten character recognition with features inspired by visual ventral stream. *Neural Processing Letters* 28(2), 97–111 (2008)

6. Bunte, K., Biehl, M., Jonkman, M.F., Petkov, N.: Learning effective color features for content based image retrieval in dermatology. *Pattern Recognition* 44(9), 1892–1902 (2011)
7. Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Convolutional Neural Network Committees for Handwritten Character Classification. In: 11th International Conference on Document Analysis and Recognition, Beijing, China, pp. 1135–1139 (2011)
8. Freitas, C.O.A., Oliveira, L.S., Aires, S.B.K., Bortolozzi, F.: Metaclasses and zoning mechanism applied to handwriting recognition. *Journal of Universal Computer Science* 14(2), 211–223 (2008)
9. Gader, P., Forester, B., Ganzberger, M., Gillies, A., Mitchell, B., Whalen, M., Yocum, T.: Recognition of handwritten using template and model-matching. *Pattern Recognition* 24(5), 421–431 (1991)
10. Hamidi, M., Borji, A.: Invariance analysis of modified C2 features: case study-handwritten digit recognition. *Machine Vision and Applications* 21(6), 969–979 (2010)
11. Hammer, B., Villmann, T.: Generalized relevance learning vector quantization. *Neural Networks* 15(8-9), 1059–1068 (2002)
12. Kauppinen, H., Seppanen, T., Pietikainen, M.: An experimental comparison of autoregressive and fourier-based descriptors in 2D shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(2), 201–207 (1995)
13. Khosravi, H., Kabir, E.: Introducing a very large dataset of handwritten Farsi digits and a study on their varieties. *Pattern Recognition Letters* 28(10), 1133–1141 (2007)
14. Klement, S., Martinetz, T.: The support feature machine for classifying with the least number of features. In: Diamantaras, K., Duch, W., Iliadis, L.S. (eds.) ICANN 2010, Part II. LNCS, vol. 6353, pp. 88–93. Springer, Heidelberg (2010)
15. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324 (1998)
16. Liu, C.L., Nakashima, K., Sako, H., Fujisawa, H.: Handwritten digit recognition: benchmarking of state-of-the-art techniques. *Pattern Recognition* 36(10), 2271–2285 (2003)
17. Oberhoff, D., Kolesnik, M.: Unsupervised shape learning in a neuromorphic hierarchy. *Pattern Recognition and Image Analysis*, 314–322 (2008)
18. Pasupathy, A., Connor, C.E.: Population coding of shape in area v4. *Nature Neuroscience* 5, 1332–1338 (2002)
19. Petkov, N.: Biologically motivated computationally intensive approaches to image pattern-recognition. *Future Generation Computer Systems* 11, 451–465 (1995)
20. Soltanzadeh, H., Rahmati, M.: Recognition of Persian handwritten digits using image profiles of multiple orientations. *Pattern Recognition Letters* 25(14), 1569–1576 (2004)
21. Wang, D., Xie, W.: Invariant image recognition by a neural networks and modified moment invariants. In: *Proceedings of the SPIE - The International Society for Optical Engineering*, SPIE; China Opt. & Optoelectron, Manuf. Assoc.; Chinese Opt. Soc. Electronic Imaging and Multimedia Systems, Beijing, China, November 4-5, pp. 217–223 (1996)
22. Wang, X.W., Ding, X.Q., Liu, C.S.: Gabor filters-based feature extraction for character recognition. *Pattern Recognition* 38(3), 369–379 (2005)
23. Chuang, Z., Zhiqing, L., Bo, X., Jun, G.: The segmentation algorithm for handwritten numeral strings in bank-check recognition. *Frontiers of Electrical and Electronic Engineering in China*, 39–44 (2007)