

# Handwritten Digit Recognition by Fourier-Packet Descriptors

Vincent Berthiaume\* and Mohamed Cheriet†

\* *Laboratory for Imagery, Vision and Artificial Intelligence, ETS, Montreal, Canada*

† *Synchromedia Laboratory for Multimedia Communication in Telepresence, ETS, Montreal, Canada*

Received 8th Nov 2011; accepted 27th Nov 2012

---

## Abstract

Any statistical pattern recognition system includes a feature extraction component. For character patterns, several feature families have been tested, such as the Fourier-Wavelet Descriptors. We are proposing here a generalization of this family: the Fourier-Packet Descriptors. We have selected sets of these features and tested them on handwritten digits: the error rate was 1.55% with a polynomial classifier for a 70 features set and 1.97% with a discriminative learning quadratic discriminant function for a 40 features set.

*Key Words:* Feature, Fourier-Wavelet Descriptor, Handwritten Digit, Rotation Invariance, Wavelet Packet.

---

## 1 Introduction

Any statistical pattern recognition system includes a *feature* extraction component. For character patterns, several feature families have been tested, such as moments and transforms [1].

The computation of some of these moments and transforms include the *Wavelet* Transform [2-7]. The Wavelet Transform performs detection of *local* components of a signal. In fact, characters contain interesting local components (extremities and breaks).

Some of these wavelet feature families have been compared to other families [2][5]: performance was better with wavelet features. That motivated us to search in this direction; we focused on the *Fourier-Wavelet Descriptors* family, by Chen and Bui [3].

The advantage of these descriptors is that not only they perform detection of local components, but they are also brilliantly conceived to be easy to be made *rotation invariant*. When reading on how their authors fixed the parameters of these descriptors, we found the next points that deserve a special attention:

- During feature extraction, the phase information of complex variables is cleared;
- Features can be slightly biased and are not perfectly invariant to rotation.
- Pixels are not all equally represented.

---

Correspondence to: <vincent.berthiaume.1@ens.etsmtl.ca>

Recommended for acceptance by <Umapada Pal>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

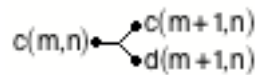


Fig. 1. Dyadic filter bank

Generally, characters are digits or letters. Digits should be easier to recognize than letters, since they are less. Therefore, before making tests on letters or letters with digits, it is preferable to start with digits alone and see if the results are promising. For the case where the digits are *handwritten*, the *Chaincode* feature family is considered by Liu and al as a state-of-the-art family [8]. With a set of these features, Liu and al. obtained, in one series of tests, error rates between 0.64% and 0.87% on a PC classifier, and between 0.68% and 1.00% on a DLQDF classifier [9].

In this paper, we propose a generalization of the Fourier-Wavelet Descriptors: the *Fourier-Packet Descriptors* (FPD). Our objective here is thus to correct the three listed points, then to test resulting FPDs sets on handwritten digits with and without rotations and then to compare the results to those obtained by Liu and al.

## 2 The Wavelet Packet Transform

A *Wavelet Function*  $\psi$  is a function having a null value at  $\pm\infty$  and having a mean value of 0. Let  $\psi_{mn}(r) \stackrel{\Delta}{=} 2^{-m/2} \psi(2^{-m}r - n)$ . The Wavelet Transform of a signal  $s$  is  $d$  in

$$d(m,n) \stackrel{\Delta}{=} \langle \psi_{mn} | s \rangle.$$

Wavelet functions are band-pass filters. The low-pass filter complementary to  $\psi$  is the *Scaling Function*  $\phi$ .

Let  $\phi_{mn}(r) \stackrel{\Delta}{=} 2^{-m/2} \phi(2^{-m}r - n)$ , then

$$c(m,n) \stackrel{\Delta}{=} \langle \phi_{mn} | s \rangle. \quad (1)$$

We see that  $d(m,n)$  and  $c(m,n)$  are correlations between  $s$  and windows  $\psi_{mn}$  and  $\phi_{mn}$ . The window  $\phi_{m,0}$  has most of its support between 0 and  $2^m$  [10]. Incrementing the variable  $n$  by 1 means sliding the window by a value of  $2^m$  to the right.

For many wavelet functions, there exist two complementary discrete filters  $g, h$  such that

$$\begin{aligned} c(m+1,n) &= \sum_k c(m,2n-k) \cdot g(k) \\ d(m+1,n) &= \sum_k c(m,2n-k) \cdot h(k) \end{aligned}$$

and such that the outputs form a complete representation of the input: it's a *Dyadic Filter Bank* [10] (Fig. 1). Equation (1) can be approximated from a finite number of samples of  $s$ : if  $m$  is negative enough, then  $2^{-m} \phi(2^{-m}r) \approx \delta(r)$ , where  $\delta$  is the Dirac Delta Function and then

$$c(m,n) \approx 2^{m/2} s(2^m(n+1)). \quad (2)$$

Several banks connected in series by the  $c$  branches form a *Dyadic Tree*. Several banks connected by any branches form a *Wavelet Packet Tree* [11]. The outputs of this tree form the coefficients of a *Wavelet Packet Transformation* of  $s$ . Therefore we see that the Wavelet Packet Transform is a generalization of the Wavelet Transform.

## 3 Selection of a Packet Transformation

It's clear that there are many different possible wavelet packet trees. For all of these trees, the global information of the outputs is the same as the input, but not distributed in the same manner between the coefficients.

When a tree is to be used in the feature extraction component of a pattern recognition system, it is desirable to select the one that maximises the inequality of the discriminant information between the output coefficients or in other words, the one for which the more *discriminant* information is concentrated in the least number of coefficients: in that way, the coefficients that have less information can be dropped with less loss of information. This operation is called the *Local Discriminant Basis* (LDB), by Coifman and Saito [12].

Let  $w_{in}^{i,k}$  and  $w_{out}^{i,k}$  be the input and output for the  $i$ th example of class  $k$ . To measure the inequality of the discriminant information between the output coefficients, the following expression can be used:

$$\sum_{n,k,l} (E_k(n) - E_l(n)) \log(E_k(n)/E_l(n)), \text{ where } E_k(n) = \sum_i w_{out}^{i,k}(n)^2 / \sum_i \|w_{in}^{i,k}\|^2. \tag{3}$$

### 4 The Fourier-Packet Descriptors Family

Before feature extraction, the pattern image can be filtered. Let  $f$  be the pattern image in polar coordinates, before filtering. Let  $\tilde{f}$  be the image in polar coordinates after size normalization and rectangular 2-D filtering in radial and angular directions i.e.

$$\tilde{f}(\rho_k, \theta_l) = \frac{1}{\alpha} \frac{\int_{\theta_{l-1}}^{\theta_l} \int_{\rho_{k-1}}^{\rho_k} f(r, t) r dr dt}{\int_{\theta_{l-1}}^{\theta_l} dt} \tag{4}$$

where  $\alpha$ , a parameter to fix, is a *size normalization* factor and where  $\rho_k, \theta_l$  are the polar coordinates of the sample  $(k, l)$  in image  $\tilde{f}$ . We see that for a varying value of  $k$  and a constant value of  $l$ , the filter sweeps a region of a form of a ring.

If the pattern in the image  $f$  is not already normalized in size, then

$$\alpha = \int_{\rho_{k-1}}^{\rho_k} r dr. \tag{5}$$

We see that this last expression multiplied by  $\pi$  is the area of the ring  $k$ . Therefore, this last equality normalizes the size of the pattern relatively to the area of the ring  $k$  so that this area becomes equal to  $\pi$  in the image  $\tilde{f}$ .

If the pattern in image  $f$  is already normalized in size and if rings are already equals in area, then

$$\alpha = 1. \tag{6}$$

Let  $R$  be the external radius of the pattern in  $f$ , then

$$\rho_k^p = \frac{k}{K} R^p \tag{7}$$

where  $p$  and  $K$  are parameters to fix. We see that  $K$  is the number of rings between the origin and the external radius. Also,

$$\theta_l = l\Delta\theta, \text{ where } \Delta\theta = 2\pi/L \tag{8}$$

and where  $L$  is a parameter to fix.

Let  $F_q$  be a coefficient of the discrete Fourier transform of  $\tilde{f}$  along the angular direction i.e.

$$F_q(\rho) = \Delta\theta \sum_{l=1}^L \tilde{f}(\rho, \theta_l) e^{jq\theta_l}. \tag{9}$$

We see that  $F_q$  has complex values, except for  $q=0$ . Let  $z$  be a parameter to fix, which is a complex variable of modulus 1 and for which  $F_q \cdot z$  is invariant to rotation. For the sake of simplicity, let

$$G_q(n) = F_q(\sqrt[n]{n}R) \cdot z. \tag{10}$$

Let's find a wavelet packet transformation of the signals  $\text{Re}(G_q)$  and  $\text{Im}(G_q)$  i.e. let  $s$  in §2 be set to

$$s = \text{Re}(G_q) \text{ or } s = \text{Im}(G_q). \tag{11}$$

Then the FPDs are defined as the coefficients of the transformation.

## 5 How the Authors Fixed the parameters

The authors use  $L < \infty$ . They use  $p = 1$ , which means that rings share the same thickness  $R/K$  (equation (7)). Also, they use  $z = F_q \left( \sqrt[q]{\sqrt[n]{nR}} \right) / F_q \left( \sqrt[q]{\sqrt[n]{nR}} \right)$  i.e.  $G_q(n) = \left| F_q \left( \sqrt[q]{\sqrt[n]{nR}} \right) \right|$  (equation (10)).

## 6 Points to Correct

In checking how the authors fixed the parameters (§5), we can notice that:

- [1] the phase of  $F_q$  is cancelled ( $G_q(n) = \left| F_q \left( \sqrt[q]{\sqrt[n]{nR}} \right) \right|$ );
- [2] rectangular filters being not ideal, the choice of  $L < \infty$  can slightly bias the Fourier coefficients (equation (9)), a phenomenon known as *Spectral Aliasing*; moreover,  $F_q \cdot z$  is invariant to rotation only for rotation values that are multiples of  $\Delta\theta$ .
- [3] all rings sharing the same thickness, the more external ones contain more pixels. Therefore, these pixels have individually less weight than pixels inside more internal rings.

## 7 Proposed Corrections

To correct these three points, our approach will consist, respectively, to:

- [1] skip the phase canceling (for the test without rotations) or replace the phase canceling by a pattern reorientation (for the test with rotations);
- [2] Choose  $L = \infty$  with  $K < \infty$  (Alternative a) or with  $K = \infty$  (Alternative b);
- [3] keep rings equal in areas i.e. choose  $p = 2$ .

Corrections 1 to 3 are detailed in the next sections 7.1 to 7.4.

### 7.1 Correction 1

Let  $\Delta$  radians be the reorientation angle i.e.  $z = e^{jq\Delta}$ . We fix  $\Delta$  so that the *principal axis* becomes vertical. For the test without rotations, no reorientation is needed:  $z = 1$ .

### 7.2 Correction 2, Alternative a

Choosing  $L = \infty$  means that  $\theta_l = \theta_{l-1}$  (equation (8)), so that equation (4) becomes

$$\bar{f}(\rho_k, t) = \frac{1}{\alpha} \int_{\rho_{k-1}}^{\rho_k} f(r, t) r dr.$$

Since  $L = \infty$ , and according to equation (8), equation (9) becomes an integral and inserting the last equation gives

$$F_q(\rho_k) = \frac{1}{\alpha} \int_0^{2\pi} \left( \int_{\rho_{k-1}}^{\rho_k} f(r, t) r dr \right) e^{jq t} dt \quad (12)$$

Let  $(x_i, y_i)$  and  $(r_i, \theta_i)$  be the  $i$ th pattern pixel in Cartesian and polar coordinates in image  $f$ . This last equation can be estimated from the discrete pattern image (Alternative a1) or from its *skeleton* (Alternative a2).

### 7.2.1 Alternative a1

Let  $\Delta A$  be the pixel area in image  $f$ . If  $\Delta A$  is small enough, then equation (12) can be approximated by

$$F_q(\rho_k) \approx \frac{1}{\alpha} \Delta A \sum_{i|\rho_{k-1} < r_i \leq \rho_k} \left( \frac{x_i + jy_i}{r_i} \right)^q. \quad (13)$$

### 7.2.2 Alternative a2

Let the pattern be reduced to its skeleton, and the pixels of the skeleton be related by segments. The pattern formed by the segments is only one dimensional. Therefore, if  $f$  is this reduced pattern and has finite values, then  $F_q$  is always 0. Therefore, to obtain values of  $F_q$  different then 0,  $f$  must be infinite on the skeleton.

Therefore, let  $f(r,t) = \delta(u(r,t))$ , where  $\delta$  is the Dirac Delta function and  $u$  is a function for which  $u(r,t) = 0$  when the point of polar coordinates  $(r,t)$  is on the skeleton. Let  $\Delta v_{ij}$  be the distance between the  $i$ th pixel and its  $j$ th neighbor, and  $n_i$  the number of neighbors of the  $i$ th pixel. If  $\Delta v_{ij}$  is small enough, then equation (12) can be approximated by

$$F_q(\rho_k) \approx \frac{1}{\alpha} \sum_{i|\rho_{k-1} < r_i \leq \rho_k} \frac{1}{n_i} \sum_j \left( \frac{x_i + jy_i}{r_i} \right)^q \Delta v_{ij}.$$

### 7.3 Correction 2, Alternative b

In Alternative  $b$ , we use assignation (5). Choosing  $L = \infty$  means that  $\theta_l = \theta_{l-1}$  (equation (8)) and  $K = \infty$  means that  $\rho_k = \rho_{k-1}$  (equation (7)), so that equation (4) with assignation (5) becomes

$$\tilde{f} = f.$$

Therefore, and since  $L = \infty$ , equation (9) becomes an integral and inserting the last equation gives

$$F_q(r) \approx \int_0^{2\pi} f(r,t) e^{jq t} dt. \quad (14)$$

This last equation can be estimated from the discrete Cartesian image. By constant interpolation, each point receives the value of its nearest pixel. On a circle, points sharing a same nearest pixel will then form an arc of constant grey level. Let  $u_i$  be the angular position of the  $i$ th arc edge and  $\hat{f}$ , the interpolated image. Since  $\hat{f}$  is constant between  $u_i$  and  $u_{i+1}$ , then

$$F_q(r) = \sum_i \int_{u_i}^{u_{i+1}} f(r,t) e^{jq t} dt \approx \sum_i \hat{f}(r, u_i^+) \cdot \begin{cases} u_{i+1} - u_i & q = 0 \\ \frac{e^{jq u_{i+1}} - e^{jq u_i}}{jq} & \text{else} \end{cases}. \quad (15)$$

### 7.4 Correction 3

Using  $p = 2$  means that rings share the same area  $\pi R^2/K$  (equation (7)).

## 8 Methodology

The last part of our objective is to test FPDs sets on handwritten digits with and without rotations and then to compare the results to those obtained by Liu and al. Let's consider the set of wavelet packet trees of input coefficients  $c(M,n)$  for values of  $n$  inside a certain interval,  $M$  being a parameter to fix. This set contains one tree for each components (real/imaginary) of  $F_q$  and for each value of  $q = 0, 1, 2, \dots$ . Our methodology consists to try progressive values of  $M$  and select the one according to a certain criteria, and then select the set of trees by LDB. To be able to compare our results to those obtained by Liu and al., the

tests must be made in the same conditions than Liu and al. We do the tests for Alternatives *a* and *b*. In the rest of this section, we specify the details of our methodology.

Characters of the same class vary in position, slant, sides ratio and size. We normalize the position of the pattern relatively to its centroid rounded to the nearest pixel corner or center, so that the coordinates of this corner or center becomes (0,0). For their series of test explained in §1, Liu and al. normalized the slant relatively to the principal axis, so that this axis becomes vertical or horizontal [9]. Therefore, for the present tests without rotations, we do the same. Now, let's consider the minimum bounding rectangle (MBR) oriented in the actual principal axis. After slant normalization, Liu and al. reduced the following ratio: maximum side of the MBR / minimum side of the MBR. Therefore, for all the present tests (with and without rotations), we do the same: we reduce this ratio to 1. Liu and al. also normalized the size relatively to the maximum side of the MBR so that this side becomes equal to a constant. We do the same for the tests of Alternative *a*: both sides becomes equal to 1. For Alternative *b*, this is not needed because the size will be normalized by the use of assignation (5). To simulate rotations, we shift the phase of Fourier coefficients.

Let's fix  $\alpha$ . For Alternative *a*, we normalize the size and the rings are of equal areas according to Correction 3 (§7.4). Therefore, we use assignation (6) for Alternative *a*. For Alternative *b*, it was already decided that we use assignation (5) (§7.3).

To estimate  $c(M,n)$ , we use approximation (2):

$$c(M,n) \approx 2^{M/2} s(2^M(n+1)) = \text{Re or Im}(G_q(2^M(n+1)))$$

according to equation (11). Let's fix the interval of  $n$  in  $c(M,n)$ , and  $K$ . The image  $f$  has a finite support: between 0 and  $R$ . Therefore, according to equations (4), (9), (10) and (11),  $s$  is supported between 0 and 1. Therefore, according to equation (1), to cover the entire signal  $s$ , the window  $\phi_{M,0}$  has to be slid a number of  $2^M - 1$  times i.e. the values of  $n$  have to be  $n = 0, 1, \dots, 2^M - 1$ . Also, we fix  $K$  to  $K = 2^M$ .

In a filtering operation, like in Fig. 1, we realize that the filters sometimes overlap the borders of the input signal. Because of this, we pad the input signal (periodically) [11] i.e.

$$c(M, n + k \cdot 2^M) = c(M, n), k \in \mathbb{Z}.$$

For the fixing of  $M$ , the criteria is the Leave-one-out error rate of a 1NN classifier. The input variables of the classifier are the 100 firsts of these:

$$c(M,0), c(M,1), c(M,2), \dots, c(M, 2^M - 1) \text{ for } s = G_0, \text{Re}(G_1), \text{Im}(G_1), \text{Re}(G_2), \text{Im}(G_2), \dots$$

$G_0$  is real and therefore  $\text{Im}(G_0)$  is skipped. The measure used for the distance between the examples in the input space is the  $L_1$  distance. We increment  $-M$  (0, 1, 2...), until the error rate stops to decrease. The greatest value of  $M$  for which the error rate decreases is the one that we select.

To select the set of trees, we extend the definition of the LDB for a *set* of trees: we redefine the  $w$  terms in expression (3) as the global input/output of *all* the trees. We choose the  $\psi$  function from the *Daubechies* family, because they are designed to have a high *regularity* [11]. When a wavelet is sufficiently regular, it can detect singularities of a certain order and below in the signal. In our signals  $\text{Re}(G_q)$  and  $\text{Im}(G_q)$ , there are discontinuities in it (singularity of order 0) and its first derivative (singularity of order 1); in the pattern, they correspond to extremities and breaks. Our choices are the first four Daubechies functions, since their regularity is just sufficient (from 0 to 1.27).

Features have to be tested on classifiers. For their series of test explained in §1, Liu and al. tested a set of 100 chaincode features on two classifiers: a polynomial classifier (PC) and a discriminative learning quadratic discriminant function (DLQDF) [9]. These classifiers included a rotation in the feature space: the PC used 70 principal components of this space and the DLQDF, 40 principal eigenvectors for each class [9]. Therefore, we perform our tests on a PC and DLQDF with 70 and 40 FPD features respectively, that we select from the 100 first output coefficients of the set of trees. For our tests, the classifiers include no rotation: the rotation is the one been performed by the set of trees (the operation done by the set of tree is orthonormal).

To select these FPD features, we use the Sequential Forward Search (SFS) method. The criteria is the Leave-one-out error rate estimator of a 1NN classifier with  $L_1$  distance.

## 9 Experimentation

For the selection of  $M$ , the LDB and the FPDs selection, we randomly chose class examples inside sections *hsf 0-1-2-3* of the *NIST SD-19* database. Let's determine  $M$ . Fig. 2 shows curves for three values of  $M$  for Alternative *a1*. We see that the curve  $-M = 3$  is below the curve  $-M = 2$  but the curve  $-M = 4$  is above the curve  $-M = 3$ . Thus, the error rate stop to decrease after  $-M = 3$  and therefore we selected the value  $-3$  for  $M$ .

Fig. 3 shows curves for Alternative *a1*, *a2* and *b* for  $-M = 3$ . Since Alternative *a1* curve is the lowest, we pursued the experiment with Alternative *a1*. Fig. 4 shows the Alternative *a1* curve for rotations simulation; we randomly fixed rotations between  $-\pi/4$  and  $\pi/4$ .

For the FPDs selection by SFS, we used 599 examples per class. For the tests on PC and DLQDF, we used the same class examples as Liu and al.: the training dataset was composed of the 66 214 examples of 600 writers (no. 0-399 and no. 2100-2299), and the test dataset was composed of the 45 398 examples of 400 writers (no. 500-699 and no. 2400-2599) [9]. The error rate was 1.55% and 1.97% for the tests on PC and DLQDF respectively.

## 10 Discussion and Interpretation

Lets try to identify the causes of misclassifications. Handwritten characters examples of the same class can vary between them by their size and tracing width. In our methodology, the scale is normalized to make the examples of identical sizes. However, it modifies the tracing width. It is thus impossible that two examples of identical tracing width but of different sizes become identical even after scale normalization. Therefore in the case of Alternative *a1* and *b*, one of the misclassifications causes can thus be the size versus the tracing width.

In the case of Alternative *a2* (skeleton), classification result was even worst than *a1* and *b* (Fig. 3). This is counter-intuitive, since the pattern is reduced to its skeleton and then the features are invariant with respect to the tracing width. Therefore, in the case of Alternative *a2*, the tracing width can't be the cause of misclassifications. A possible cause could be that the FPDs are more sensitive to intraclass variation when they are estimated from the skeleton.

On the test with rotations, classification results were less good than without rotations (Fig. 4). The cause for this can be the possible confusion between '6' and '9'.

## 11 Conclusion

Our objective was to correct the three already clarified points (§6), then to test resulting FPDs sets on handwritten digits with and without rotations and then to compare the results to those obtained by Liu and al. We effectively have brought a correction to each point (§7).

Classification results were better for Alternative *a1* than for Alternative *a2* and *b* (Fig. 3), so that we pursued the experiment with Alternative *a1*. Rotations simulation gave the result on Fig. 4.

Tests of the FPDs sets gave an error rate of 1.55% with PC and 1.97% with DLQDF. We see that these results are not as good as Liu and al. (between 0.64% and 0.87% and between 0.68% and 1.00%).

Let's see how we could improve the results. For Alternative *a1*, to estimate  $F_q(\rho_k)$  from the discrete pattern image, we used a weighted sum of the pattern pixels (equation (13)). To have a better estimation of  $F_q(\rho_k)$ , we could insert new pixels by bilinear interpolation. For Alternative *b*, to estimate equation (14) from the discrete image, we used constant interpolation (equation (15)). Instead, we could use bilinear interpolation. Once we'll have improved the results for handwritten digits, we could test FPDs on handwritten letters and on handwritten letters *and* digits.

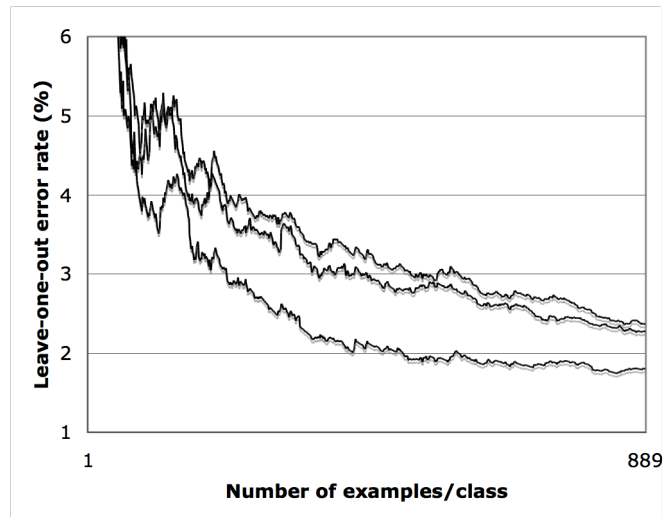


Fig. 2: Curves -M = 2, 4, 3 (top to bottom). Alternative *a1*.

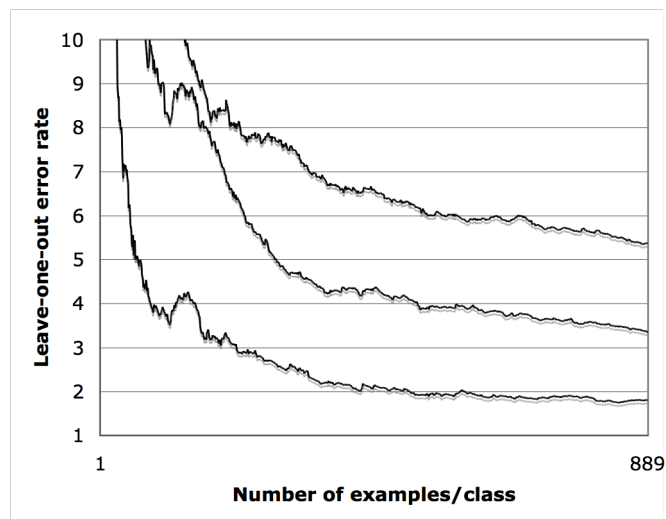


Fig. 3: Curves for Alternative *a1*, *b* and *a2* (bottom to top).

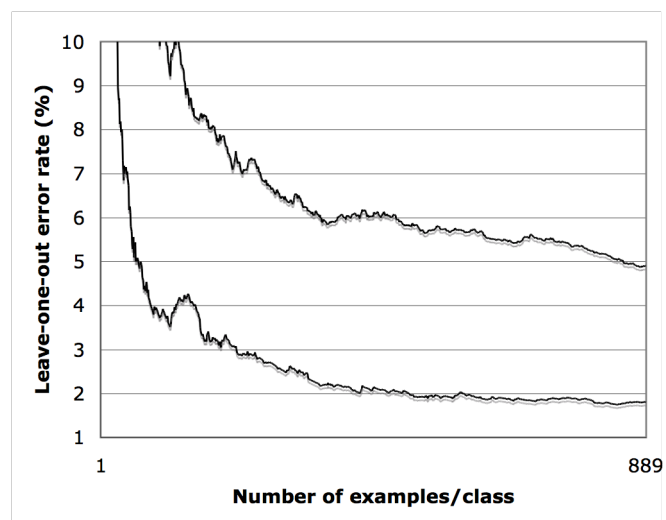


Fig. 4. Curve for Alternative *a1* with rotations (top) and without rotations (bottom).



## Acknowledgement

We'd like to thank Cheng-Lin Liu, Professor from National Laboratory of Pattern Recognition, for the PC and DLQDF tests. We'd also like to thank Mathias Adankon, searcher at Laboratory for Imagery, Vision and Artificial Intelligence, for an SVM test, which we have not retained in this paper.

## References

- [1] O.D. Trier, A.K. Jain and T. Taxt, "Feature extraction methods for character recognition—a survey," *Pattern Recognition*, vol. 29, no. 4, pp. 641–662, April 1996.
- [2] D. Shen and H. H. S. Ip, "Discriminative wavelet shape descriptors for description of 2-d patterns," *Pattern Recognition*, vol. 32, no. 2, pp. 151-165, February 1999.
- [3] G. Chen and T.D. Bui, "Invariant fourier-wavelet descriptor for pattern recognition," *Pattern Recognition*, vol. 32, no. 6, pp. 1083-1088, June 1999.
- [4] Y. Y. Tang, B. F. Li, H. Ma and J. Liu, "Ring-projection-wavelet-fractal signatures: a novel approach to feature extraction," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 45, no. 8, pp. 1130-1134, August 1998.
- [5] P. Wunsch and A.F. Laine, "Wavelet descriptors for multiresolution recognition of handprinted characters," *Pattern Recognition*, vol. 28, no. 8, pp. 1237-1249, August 1995.
- [6] S.-W. Lee, C.-H. Kim, H. Ma, Y.Y. Tang, "Multiresolution recognition of unconstrained handwritten numerals with wavelet transform and multilayer cluster neural network," *Pattern Recognition*, vol. 29, no. 12, pp. 1953-1961, December 1996.
- [7] A. F. Laine and S. Schuler, "Hexagonal wavelet representations for recognizing complex annotations," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 740-745, 1994.
- [8] C.L. Liu, K. Nakashima, H. Sako and H. Fujisawa, "Handwritten digit recognition: benchmarking of state-of-the-art techniques," *Pattern Recognition*, vol. 36, no. 10, pp. 2271–2285, October 2003.
- [9] C.L. Liu, K. Nakashima, H. Sako and H. Fujisawa, "Handwritten digit recognition: investigation of normalization and feature extraction techniques," *Pattern Recognition*, vol. 37, no. 2, pp. 265–279, February 2004.
- [10] Vetterli, Martin and Kovacevic, Jelena, *Wavelets and subband coding*. Englewood Cliffs, N.J.: Prentice-Hall, 1995.
- [11] The Mathworks, *Wavelet Toolbox user's guide*, [mathworks.com/access/helpdesk/help/toolbox/wavelet](http://mathworks.com/access/helpdesk/help/toolbox/wavelet)
- [12] R. R. Coifman and N. Saito, "Constructions of local orthonormal bases for classification and regression," *Comptes Rendus Acad. Sci. Paris, Serie I*, no. 319, pp.191-196, July 1994.