A Wavelet-based Descriptor for Handwritten Numeral Classification

Leticia M. Seijas and Enrique C. Segura
Departamento de Computación, Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires
Buenos Aires, Argentina
{*lseijas, esegura*}*@dc.uba.ar*

***Abstract*—In this work we propose descriptors for handwritten digit recognition based on multiresolution features by using the CDF 9/7 Wavelet Transform and Principal Component Analysis, in order to improve the classification performance and obtain a strong reduction on the size of the digit representation. This allows for a higher precision in the recognizers and, at the same time, lower training costs, especially for large datasets. Experiments were carried out with the CENPARMI and MNIST databases, widely used in the literature for this kind of problems, combining classifiers of the Support Vector Machine type. The recognition rates are good, comparable to those reported in previous works.**

***Keywords*-multiresolution features; digit recognition; Support Vector Machines; dimension reduction; descriptor**

I. Introduction

The Wavelet Transform (WT) is a technique particularly suited for locating spatial and frequential information in image processing and, particularly, for feature extraction from patterns to be classified. Many works have applied WT in different areas [1] [2] [3]. For the case of handwritten digit recognition, there are several approaches. In [4] a discrete, one-dimensional orthogonal wavelet is applied to the digit contour and a descriptor is constructed with the approximation bands of the transform; here the aim is to prevent variations of writing style from affecting the classification. [5] introduces a descriptor that applies orthonormal wavelets on the normalized contour of each digit, up to third detail level, so as to obtain a smoothed pattern representation. In [6] the discretized continuous wavelet transform Mexican Hat is applied in order to produce a smaller version of each digit and the Wavelet Gradient is used in order to construct an additional vector with orientation, gradient and curvature features at different scales. [7] uses biorthogonal wavelets *Cohen-Daubechies-Feauveau (CDF)*, in its two-dimensional form, obtaining a descriptor with all the four first-level subbands, normalized. Preprocessing is a fundamental step in classification, since it determines which features will be relevant for discriminating patterns of different classes. Dimensionality reduction has a strong incidence in performance and computational cost; frequently it enables application of certain computational techniques that otherwise would become impractical or impracticable. This work presents a wavelet-based descriptor for handwritten digit classification that allows a strong reduction of dimensionality while increasing the quality of the pattern representation and the recognition rates. The applied techniques for preprocessing are well suited for the problem, nevertheless, an appropriate combination of them has resulted in an interesting proposal. Note that the objective of this work does not include the description of the system recognizer, presented in [8]. Finally, since this work is focused on correct pattern classification, we are interested in pattern representation rather than their reconstruction from the WT coefficients. Section II gives an overview of the Discrete Wavelet Transform. In Section III descriptors for handwritten digits are proposed. Experiments for testing the proposed descriptors with different recognition systems are reported in Section IV and, finally, in Section V we present the conclusions of the work.

II. Discrete Wavelet Transform

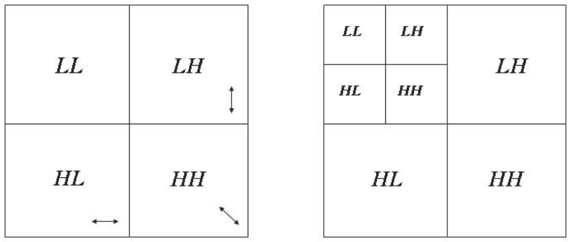
The Discret Wavelet Transform (DWT) is based on the subband-coding technique, being a variant easy to implement of WT that requires a few resources and computing time. The DWT is well suited for multiresolution analysis (MRA) and lossless reconstruction by the use of filter banks [9].

The Fast Orthogonal Wavelet Transform (FWT) decomposes each approximation $P_{V_j} f$ of the function $f \in L^2(\Re)$, into approximations of lower resolution $P_{V_{j+1}} f$ plus wavelet coefficients produced by the projection $P_{W_{j+1}} f$, being $V_{j\, j \in Z}$ a multiresolution approximation [10], W_j the orthogonal complement of V_j and $P_{V_j} f$ the orthogonal projection of f in V_j. Conversely, for the reconstruction from wavelet coefficients, each $P_{V_j} f$ is obtained through $P_{V_{j+1}} f$ and $P_{W_{j+1}} f$. Besides, $\{\varphi_{j,n}; j, n \in Z\}$ and $\{\psi_{j,n}; j, n \in Z\}$ are orthonormal bases for V_j and W_j, with φ and ψ being scale and wavelet functions respectively.

Digital image processing requires a two-dimensional WT. It is computed by application of the one-dimensional FWT first to rows and then to columns. Let $\psi(x)$ be the one-dimensional wavelet associated to the one-dimensional scale function $\varphi(x)$, then the scale function in two dimensions is:

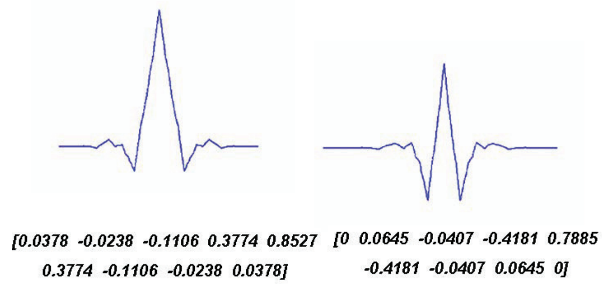
$$\varphi(x, y)_{LL} = \varphi(x)\varphi(y) \tag{1}$$

978-0-7695-4774-9/12 $26.00 © 2012 IEEE
DOI 10.1109/ICFHR.2012.174



(a) First level (b) Second level

Figure 1. Multilevel decomposition of an N x N image using a 2D-DWT.



(a) Scaling function (b) Wavelet function

Figure 2. CDF 9/7 with filters for signal decomposition.

and the three two-dimensional wavelets are defined by

$$\psi(x,y)_{LH} = \varphi(x)\psi(y) \tag{2}$$

$$\psi(x,y)_{HL} = \psi(x)\varphi(y) \tag{3}$$

$$\psi(x,y)_{HH} = \psi(x)\psi(y) \tag{4}$$

where LL represents lowest frequencies (global information), LH represents high vertical frequencies (horizontal details), HL high horizontal frequencies (vertical details) and HH high frequencies on both diagonals (diagonal details).

Application of a step of the transform on the original image, produces an approximation subband LL corresponding to the smoothed image, and three detail subbands HL, LH and HH. The following step works on the approximation subband, resulting in four subbands (see Fig. 1). In other words, each step of the decomposition represents the approximation subband of level i as four subbands at level $i+1$, each one being a quarter in size respecting the original subband.

Although the standard DWT is a powerful technique, it shows some limitations. For example, it is sensitive to translations and only detects a few directional features. Nevertheless, we believe that the use of this type of wavelet for handwritten digit recognition is reasonable, since the images that we are dealing with are normalized in size and centered, and the directionality of a digit, even handwritten, does not present substantial variations.

III. Descriptors for Handwritten Numerals

A. Application of the Wavelet Transform

The Biorthogonal Two-dimensional Wavelet *Cohen-Daubechies-Feauveau (CDF) 9/7* is particularly suited for our research line [8]. It is efficiently applied e.g. to the JPEG2000 compression standard, and also in fingerprint compression by the FBI [11]. Fig. 2 shows the scale and wavelet functions and the coefficients of the corresponding filters, for the CDF 9/7 transform in the decomposition.

Our experiments were carried out with the CENPARMI and MNIST handwritten digit databases, both widely used in the literature. The CENPARMI database [12] consists of digitized images in two levels, that we have normalized to a size of 16x16 pixels [13]. It includes training and testing sets, with 4000 and 2000 labeled patterns respectively. The MNIST database [14] contains 70000 256-level grayscale images, labeled as 60000 for training and 10000 for testing, size 28x28.

Fig. 3 shows the application of the CDF 9/7 up to a second level on CENPARMI and MNIST patterns. It can be seen that in the second level some details of the digits are lost, hence we decided, by the moment, not to consider further levels of the transform.

Different combinations were evaluated for constructing the descriptor: first and second level approximation bands (LL) produce a smoothed image of the pattern, preserving shape and reducing dimension to a quarter of the original in a first level, and to 16th in a second level, where the image is coarser. The high frequency subband HH of first level, shows sudden changes in image contours (diagonal details), that we think might contribute in detecting differences between patterns, and also the LH and HL subbands. The second level detail subbands provide high frequency features -vertical, horizontal and diagonal- on the smoothed digit, which can show basic structural features of the pattern, useful in the classification process.

Another issue to be considered is what values are representative in the obtained subbands. Therefore some descriptors were thresholded and binarized. To do this, several statistics were computed over the values of each subband,

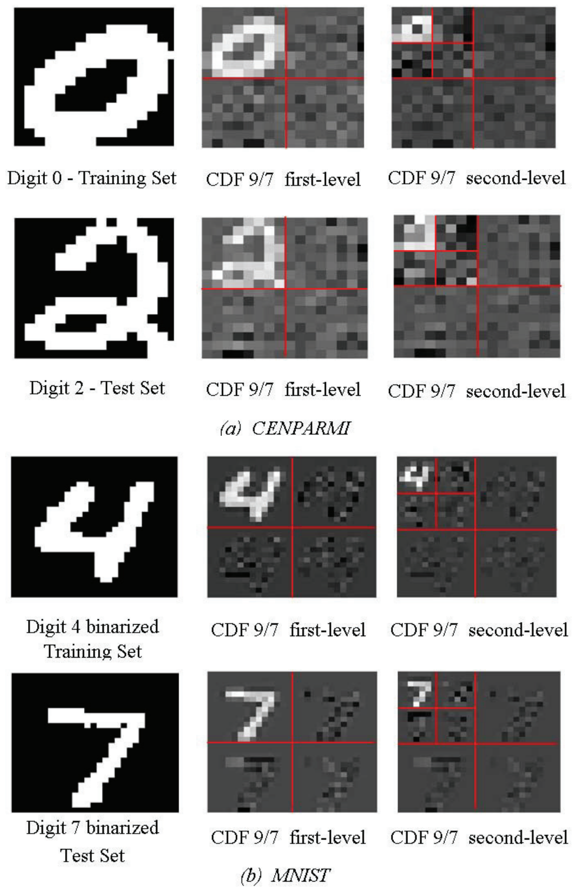


Figure 3. Applying wavelet CDF 9/7 to handwritten digits extracted from standard databases.

such as mean, median, variance and standard deviation.

For MNIST we extracted 15000 patterns from the training set, while the test set was used in its entirety.

For testing the descriptors we have applied Support Vector Machines (SVM), a well-known and efficient classification technique [15] [16] [17].

The best results correspond to non-thresholded representation: first level approximation subband (LL1), all four second level subbands (T2) and the combination of these two descriptors (LL1T2), for both databases.

Tables I and II show the best results for SVM classifiers.

Table I
RECOGNITION OF *CENPARMI* TEST SET FOR LL1, T2 AND LL1T2 DESCRIPTORS USING GAUSSIAN SVMS WITH PARAMETER $\sigma = 5.50$, 4.75 AND 5.60 RESPECTIVELY.

Descriptor	Dimension	SVM (% Recognied)
Initial	256	85.00
LL1	64	94.20
T2	64	94.45
LL1T2	128	*94.75*

Table II
RECOGNITION OF *MNIST* TEST SET FOR LL1, T2 AND LL1T2 DESCRIPTORS USING GAUSSIAN SVMS WITH PARAMETER $\sigma = 15.00$, 15.00 AND 21.00 RESPECTIVELY (CLASSIFIERS TRAINED WITH 15000 BINARIZED IMAGES EXTRACTED FROM MNIST TRAINING SET).

Descriptor	Dimension	SVM (% Recognized)
Binarized	784	97.33
LL1	196	97.59
T2	196	97.54
LL1T2	392	*97.60*

All four second level subbands of the CDF 9/7 represent the digits efficiently, with a dimension reduction of 75%. Results improve when incorporating the first level approximation, where the smoothed image is less coarse than that of second level. We believe that the first level approximation makes a contribution regarding the basic structure of the digits, and this is reflected by the results. In this case the dimension reduction is 50%. On the other hand, using the first level approximation subband as a descriptor is also a good choice, as it shows a good performance and reduces dimension in 75%.

Thus, the application of the WT up to a second level of resolution is a suitable representation of the digits by using approximation and detail subbands and is associated with a considerably reduction of the size descriptor. Although it is not the main objective of the present work, definition of a general methodology for deciding the number of levels of WT to apply and which subbands to choose, could be considered for the future.

B. Application of PCA to Descriptors based on WT (WT-PCA)

Reduction of descriptor size is important, especially for large bases with high-dimensional data, as in the case of MNIST, whose patterns are 28x28, and therefore would require a descriptor of 784 (three times the CENPARMI images). In addition, MNIST has 70000 patterns (almost twelve times the size of CENPARMI).

Principal Component Analysis (PCA) is useful for finding the directions that contain most of the data variance, disregarding those with little information [18]. This allows to reduce dimension and computational costs and, at the same time, to maintain or possibly increase the quality of classification. Sometimes this point is crucial to decide if a certain technique is applicable, depending on the availability of computational time and resources.

PCA was applied to the WT-based descriptors with best performance in previous Subsection: LL1, T2 and LL1T2, i.e. first level approximation, second level transform, and the descriptor formed with both, using the CDF 9/7.

There is no a priori rule to decide the optimal number of principal components to use. We were interested in drastically reducing dimension while retaining most of the

Table III
RECOGNITION OF *CENPARMI* TEST SET FOR DESCRIPTORS BASED ON DIRECTIONAL FEATURES AND WT-PCA, USING SVMS (HR - HORIZONTAL, VT - VERTICAL, RD - RIGHT DIAGONAL, LD - LEFT DIAGONAL, GL - GLOBAL).

	Descriptor	LL1 (dim.64)	T2 (dim.64)	LL1T2 PCA 64
	HR	90.10	89.70	89.80
	VT	91.35	91.25	91.40
%	RD	91.60	91.10	90.95
	LD	93.00	92.55	92.65
	GL	94.60	94.35	94.50

Table IV
RECOGNITION OF *MNIST* TEST SET FOR DESCRIPTORS BASED ON DIRECTIONAL FEATURES AND WT-PCA, USING SVMS (HR - HORIZONTAL, VT - VERTICAL, RD - RIGHT DIAGONAL, LD - LEFT DIAGONAL, GL - GLOBAL). THE CLASSIFIERS WERE TRAINED WITH THE COMPLETE TRAINING SET.

	Descriptor	LL1 PCA 98	T2 PCA 98	LL1T2 PCA 196
	HR	97.35	97.27	97.30
	VT	97.67	97.60	97.66
%	RD	97.65	97.65	97.61
	LD	97.68	97.68	97.70
	GL	98.64	98.54	98.59

Table V
RECOGNITION OF *CENPARMI* TEST SET USING THE STRATEGY OF COMBINING DIRECTIONAL AND GLOBAL FEATURES.

Descriptor	% Total (reconocidos)	% Error
LL1	97.35	2.65
T2	97.25	2.75
LL1T2 PCA 64	*97.40*	*2.60*

Table VI
RECOGNITION OF *MNIST* TEST SET USING THE STRATEGY OF COMBINING DIRECTIONAL AND GLOBAL FEATURES.

Descriptor	% Total (reconocidos)	% Error
LL1 PCA 98	99.29	0.71
T2 PCA 98	*99.32*	*0.68*
LL1T2 PCA 196	99.27	0.73

information. After several experiments varying the number of principal components, it was possible to retain more than 90% of the total variance while reducing the descriptor size in at least 50% for best recognition rates using SVMs. We note that for low dimension descriptors (such as LL1 and T2 for CENPARMI), PCA did not improve results. In the case of LL1T2 represented with 64 components (LL1T2 PCA 64), a similar performance was obtained to that with wavelet descriptor without PCA, even with reduction of dimension to a half. For MNIST descriptors, applying PCA not only reduces dimensionality but also increases recognition percentages, being the best result with LL1 PCA 98, T2 PCA 98 and LL1T2 PCA 196 (98.04%, 97.94% and 97.96% respectively, for a subset of 15000 training patterns).

In this way, we select descriptors LL1, T2 and LL1T2 PCA 64 for CENPARMI, reducing dimension in 75.00%. For MNIST we choose descriptors LL1 PCA 98, T2 PCA 98 and LL1T2 PCA 196, reducing dimension in 87.50% in the two former cases and in 75.00% in the latter, with good classification percentages.

IV. CLASSIFYING WITH WT-PCA DESCRIPTORS

The descriptors introduced in the previous Section were applied to constructing and testing handwritten digit recognition systems based on combination of individual classifiers. We worked with four directional features horizontal, vertical, right and left diagonal extracted from the digits by the use of Kirsch masks, and on them were applied WT-PCA descriptors. The Kirsch edge detector is a standard for handwritten numeral representation, considered as precise and robust as compared to other techniques [19] [20]. Besides the WT-PCA descriptors on the directional features, their application to the whole pattern was considered, which we call 'global feature'. It is reported that SVMs using Gaussian kernel outperform traditional techniques for the handwritten digit recognition problem [17] [19]. So, we used them as individual classifiers obtaining high recognition rates (see Tables III and IV).

The Bayesian strategy with ambiguous pattern detection we have presented in [8], was used for the combination of classifiers. This strategy provided high classification percentages. So, the recognition system was composed of two levels. The first one was formed by a collection of independent classifiers, each one specialized in a different feature extracted from the input pattern. The second level consisted of an analyzing module in charge of defining and explaining the output of the system.

Tables V and VI show the results for both databases. Each recognition system combines all four directional features and the global one, for a particular WT-PCA descriptor [8] [21] [22].

For CENPARMI, the highest recognition rate was *97.40%*, achieved with the *LL1T2 PCA 64* descriptor. For MNIST the highest percentage was *99.32%*, obtained with the *T2 PCA 98* descriptor for the system with classifiers trained with the whole dataset. These results improve previous ones reported in the literature [8] [21], both in recognition rates and in complexity of the classifiers, due to dimensionality reduction.

V. CONCLUSION

In this paper we proposed descriptors for handwritten numeral recognition based on multiresolution features by the use of the CDF 9/7 Wavelet Transform and Principal Component Analysis (WT-PCA), which improved classification performance and considerably reduced dimensionality of the representation. The applied techniques for preprocessing are known and well suited for the problem, nevertheless an appropriate combination of them resulted in an interesting proposal, being the main contribution of this work.

Experiments were performed on the CENPARMI and MNIST databases, widely accepted as benchmarks for this problem. Five classifiers of SVM type were combined with the Bayesian technique that detects ambiguous patterns, introduced in [8]. Each individual classifier was dedicated to a specific directional or global feature, to which a descriptor WT-PCA was applied.

For CENPARMI the best result was obtained with the *LL1T2 PCA 64* descriptor associated with *97.40%* of correct classification, reducing input dimension in 75%. For MNIST the highest percentage was achieved with *T2 PCA 98*: *99.32%*, lowering input dimension in 87.50%. Reduction of descriptor size has a decisive impact on training time and on processing of large databases.

These recognition percentages are high, improving rates reported in several previous works [23] [24] [20] [5] [7]. On the other hand, approaches with better results than ours mostly use modified training sets with additional images obtained by applying deformations on patterns of the training set [25] [26], and increasing considerably the number of training images. According to [27] the best results for MNIST were achieved with that technique, which suggests the possibility, as a future work, of incorporating it to our proposal.

Acknowledgment

This work has been partially supported by projects UBACYT X199 from the Universidad de Buenos Aires and PICT 26001 from Argentine government.

References

[1] P.Nicholl, A. Amira, D. Bouchaffra, and R. Perrott, “A statistical multirresolution approach for face recognition using structural hidden markov models,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–13, 2008.

[2] L. Kaplan and R. Murenzi, “Pose estimation of sar imagery using the two dimensional continuous wavelet transform,” *Pattern Recognition Letters*, vol. 24, pp. 2269–2280, 2003.

[3] C. Chen, C. Chen, and C. Chen, “A comparison of texture features based on svm and som,” 18th International Conference on Pattern Recognition. Hong Kong: ICPR 2006, 2006, pp. 630–633.

[4] P. Wunsch and A. Laine, “Wavelet descriptors for multiresolution recognition of handprinted characters,” *Pattern Recognition*, vol. 28, no. 8, pp. 1237–1249, 1995.

[5] G. Y. Chen, T. D. Bui, and A. Krzyzak, “Contour-based handwritten numeral recognition using multiwavelets and neural networks,” *Pattern Recognition*, vol. 36, no. 7, pp. 1597–1604, 2003.

[6] D. Romero, A. Ruedin, and L. Seijas, “Wavelet based feature extraction for handwritten numerals,” *Image Analysis and Processing (ICIAP 2009), LNCS 5716, Springer*, pp. 374–383, 2009.

[7] S. Correia and J. de Carvalho, “Optimizing the recognition rates of unconstrained handwritten numerals using biorthogonal spline wavelets,” Pattern Recognition, 2000. Proceedings. 15th International Conference on. Barcelona, Spain: ICPR 2000, 2000, pp. 251–254.

[8] L. Seijas and E. Segura, “Detection of ambiguous patterns using svms: Application to handwritten numeral recognition,” *Computer Analysis of Images and Patterns (CAIP 2009) - Lecture Notes in Computer Science, LNCS 5702*, pp. 840–847, 2009.

[9] P. Shukla, “Complex wavelet transforms and their applications,” M.Phil. Thesis. Glasgow (United Kingdom): Dept.of Electronic and Electrical Engineering, University of Strathclyde, 2003.

[10] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[11] A. Skodras, C. Christopoulos, and T. Ebrahimi, “JPEG2000: The upcoming still image compression standard,” *Elsevier, Pattern Recognition Letters*, vol. 22, pp. 1337–1345, 2001.

[12] C. Suen, C.Nadal, R. Legault, T. Mai, and L. Lam, “Computer recognition of unconstrained handwritten numerals,” *Procs IEEE*, vol. 80, no. 7, pp. 1162–1180, 1992.

[13] S. B. Cho, “Ensemble of structure-adaptive selforganizing maps for high performance classification,” *Information Sciences*, vol. 123, pp. 103–114, 2000.

[14] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[15] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, New York, 1995.

[16] C. Liu and H. Fujisawa, “Classification and learning methods for character recognition: Advances and remaining problems,” in *Machine Learning in Document Analysis and Recognition*, H. F. S. Marinai, Ed. Springer, 2008, pp. 139–161.

[17] L. Oliveira and R. Sabourin, “Support vector machines for handwritten numerical string recognition,” 9th IEEE International Workshop on Frontiers in Handwritten Recognition. Washington DC: IEEE Computer Society, 2004, pp. 39–44.

[18] S. Haykin, *Neural Networks A Comprehensive Foundation*. Prentice Hall, 1999.

[19] C. Liu, K.Nakashima, H. Sako, and H. Fujisawa, “Handwritten digit recognition: benchmarking of state-of-the-art techniques,” *Pattern Recognition*, vol. 36, pp. 2271–2285, 2003.

[20] D. Gorgevik and D. Cakmakov, “An efficient three-stage classifier for handwritten digit recognition,” vol. 4, 17th International Conference on Pattern Recognition. IEEE Computer Society, 2004, pp. 507–510.

[21] L. Seijas and E. Segura, "Detection of ambiguous patterns in a som based recognition system: application to handwritten numeral classification," 6th International Workshop on Self-Organizing Maps. Germany: Bielefeld University, 2007.

[22] S. Lee, "Off-line recognition of totally unconstrained handwritten numerals using multilayer cluster neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 648–652, 1996.

[23] B. Kegl and R. Busa-Fekete, "Boosting products of base classifiers." Montreal, Canada: Proceedings of the 26th International Conference on Machine Learning, 2009.

[24] F. Lauer, C. Suen, and G. Bloch, "A trainable feature extractor for handwritten digit recognition," *Pattern Recognition*, vol. 40, pp. 1816–1824, 2007.

[25] C. Suen, K. Kiu, and N. Strathy, "Sorting and recognizing cheques and financial documents," *Document Analysis Systems: Theory and Practice, S.W.Lee and Y.Nakano (Eds.), LNCS 1655, Springer*, pp. 173–187, 1999.

[26] C. Liu and H. Fujisawa, "Classification and learning for character recognition: Comparison of methods and remaining problems." Seoul: International Workshop on Neural Networks and Learning in Document Analysis and Recognition, 2005.

[27] D. Ciresan, U. Meier, L. Gambardella, and J. Schmidhuber, "Handwritten digit recognition with a committee of deep neural nets on gpus," *Technical Report No. IDSIA-03-11. IDSIA / USI-SUPSI, Dalle Molle Institute for Artificial Intelligence, Switzerland*, 2011.