Towards a SignWriting Recognition System

D. Stiehl, L. Addams, L. S. Oliveira
UFPR, DInf, PPGInf
Curitiba, PR, Brazil
Email: lesoliveira@inf.ufpr.br

C. Guimarães
UTFPR
Curitiba, PR, Brazil
Email: cayleyg@utfpr.edu.br

A. S. Britto Jr.
PUCPR, PPGIa
Curitiba, PR, Brazil
Email: alceu@ppgia.pucpr.br

***Abstract*—SignWriting is a writing system for sign languages. It is based on visual symbols to represent the hand shapes, movements and facial expressions, among other elements. It has been adopted by more than 40 countries, but to ensure the social integration of the deaf community, writing systems based on sign languages should be properly incorporated into the Information Technology. This article reports our first efforts toward the implementation of an automatic reading system for SignWiring. This would allow converting the SignWriting script into text so that one can store, retrieve, and index information in an efficient way. In order to make this work possible, we have been collecting a database of hand configurations, which at the present moment sums up to 7,994 images divided into 103 classes of symbols. To classify such symbols, we have performed a comprehensive set of experiments using different features, classifiers, and combination strategies. The best result, 94.4% of recognition rate, was achieved by a Convolutional Neural Network.**

I. Introduction

Writing systems are maybe the most important invention of the mankind, and play a major role in the modern society. They are used for literature, cultural preservation, information storage and retrieval, science, knowledge creation, communication among many others vital societal functions [5]. However, deaf people have difficulties to acquire a writing system of oral language which is mostly based on phonemes. The deaf communicate through Sign Language, their natural language, of visual-spatial manner. The structure of a Sign Language is completely different from the sequential frame of a written/spoken language, due to the use of multi-linear (both spatial and temporal) relationships among signs and their components.

Valerie Sutton [12] proposed a writing system called SignWriting which uses visual symbols to represent the hand shapes, movements, facial expressions and other features of the signed languages. It is considered an alphabet, i.e., a list of symbols used to write any signed language in the word. Therefore, SignWriting makes it possible to have books, newspapers, magazines, dictionaries, daily notes and messages and literature written in signs. It can be used to teach signs and signed language grammar to beginning signers, or it can be used to teach other subjects, such as math, history, or English to skilled signers. According to Bianchini et al [1], SignWriting is a very promising solution for Sign Language transcription, compared to other proposed writing systems.

The alphabet of the SignWriting is composed of seven categories of base symbols: Hand, Movement, Dynamics & Timing, Head & Face, Body, Detailed Location, and Punctuation. The signs of the sign language are written using these symbols. Figure 1 shows three different words written in Sign Language.

The sign in American Sign Language equivalent to the word sorrow is represented by both hands facing the signer, with all fingers open and extended, in a simultaneous downward movement. Notice that some symbols may overlap each other, e.g., hands and face. In the second sample, foolish is represented by the face with eyebrows and mouth contracted and hands sideways in relation to the signer with index and middle fingers open and extended moving towards each other twice. Finally, the third example, pride, is represented by one hand closed, with thumb open and extended, sideways in relation to the signer, in an spiral upwards movement.

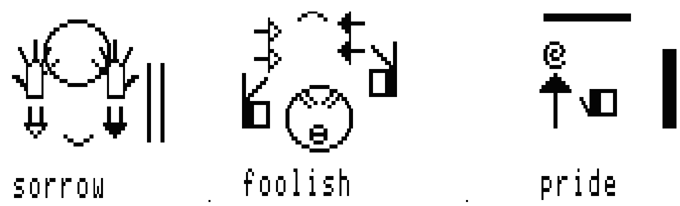


Fig. 1. Example of three worlds written in SignWriting: (a) sorrow, (b) foolish, and (c) pride

According to Guimares et al. [5], SignWriting has been adopted by more than 40 countries, but to ensure the social integration of the deaf community, writing systems based on sign languages should be properly incorporated into the Information Technology [4]. In this regard, two issues must be addressed to facilitate and popularize the use of this writing system. First, there is a lack of an efficient SignWriting editor so that people can easily take notes and disseminate the knowledge. The ones of which we are aware are based on a series of menus where one must find the symbols and drag them to an specific area to form the desired written sign. This procedure is time consuming and unstimulating. Figure 2 shows an example of such an interface [1].

A second issue is the lack of an engine to convert the SignWriting script into text so that one can store, retrieve, and index information in an efficient way. Such an engine would be similar to an OCR (Optical Character Recognition), but instead of recognising characters, it would recognise the glyphs used by the SignWriting, hence, an OGR (Optical Glyph Recognition). From Figure 1 we can see that building an OGR is not a straightforward task due to the great number of glyphs, the overlap that occurs among them and the intrinsic variability of the handwriting. The challenges of this kind of application are discussed from different perspectives in [3] and [13].

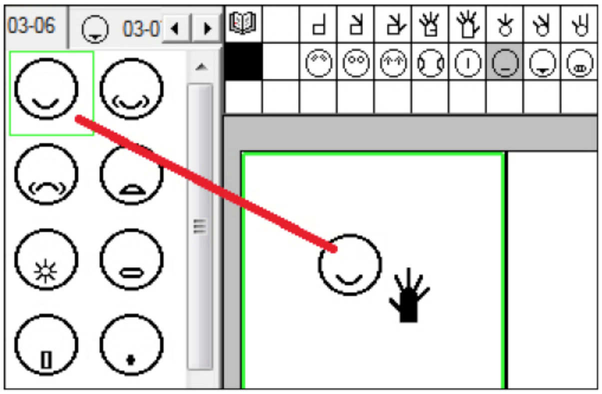


Fig. 2. User interface of a classical SignWriting editor

In this paper we report our first efforts toward the implementation of an OGR for the Brazilian Signed Language (LIBRAS). To the best of our knowledge, there is no database available for research on SignWriting recognition. To overcome this difficulty, we developed an interface to collect such data that can be used to foment the research in the field of SignWriting recognition and make future benchmark and evaluation possible. Currently, the data contains about 8,000 images of 103 different hand configurations and it is available for research purposes under request[1]. More details about the procedures used to collect the database can be found in [6]. To gain better insight into this dataset, we tested the state-of-the-art representations and classifiers used for handwriting recognition. The best result, 94.4% of recognition rate, was achieved by using Convolutional Neural Network with data augmentation.

II. SignWriting

Widely used, SignWriting is becoming the standard writing system for sign languages. The International SignWriting Alphabet (ISWA) proposed in 2010 includes all symbols used to write the handshapes, movements, facial expressions, and body gestures of most sign language in the world.

SignWriting is divided into 7 categories and 30 groups of symbols, which are summarised in Table I. The category 1, "Hands", contains 261 symbols distributed into 10 groups. The second category, "Movement" contains 242 contact symbols, small finger movements, straight arrows, curved arrows and circles, which are placed into 10 groups based on planes. The third category, "Dynamics and Timing" is composed of 8 symbols that are used mostly with movement symbols and punctuation symbols to give the feeling or tempo to movement. They also provide emphasis on a movement or expression, and combined with Puncuation Symbols become the equivalent to Exclamation Points. The Tension Symbol, combined with Contact Symbols, provides the feeling of "pressure", and combined with facial expressions can place emphasis or added feeling to an expression. Timing symbols are used to show alternating or simultaneous movements.

The forth category, "Head & Face" includes 114 symbols to describe the head movement and the positions of the head.

[1]http://web.inf.ufpr.br/vri/signwriting-database

TABLE I. Categories and Groups of Symbols used by the SignWriting

Category	Number of Symbols	Number of Groups	Description
Hands	261	10	Index, Index middle, Index middle thumb, four fingers, five fingers,baby finger, ringer finger, middle finger, index thumb, thumb
Movement	242	10	Contact, Finger movement, Straight Wall Plane, Straight Diagonal Plane, Straight Floor Plane, Curves Hit Wall Plane, Curves Parallel Wall Plane, Curves Hit Floor Plane, Curves Parallel Floor Plane, Circles
Dynamics & Timing	8	1	Dynamics & Timing
Head & Face	114	5	Head, Brow Eyes Eyegaze, Cheeks Ears Nose Breath, Mouth Lips, Tongue Teeth Chin Neck
Body	18	2	Trunk, Limbs
Location	8	1	Location
Punctuation	5	1	Punctuation

Some groups contain detailed facial expressions and movement of parts of the face and neck. Category 5, "Body" is composed of 6 symbols representing torso movement, shoulders, hips, and the limbs. This parts of the body are used in Sign Languages as a part of grammar, especially when describing conversations between people, called Role Shifting, or making spatial comparisons between items on the left and items on the right. The sixth category, "Location", contains 8 symbols but are not used when writing signs on a daily basis. The symbols of this category are only used in computer software to assist in giving further details for sorting large sign language dictionaries that are sorted by SignWriting symbols. Finally, category 7, "Punctuation", contains 5 symbols that are used when writing complete sentences or documents in SignWriting. The Punctuation Symbols do not look like the symbols for punctuation in English, but they do have similar meanings. Figure 3 shows some examples of the symbols of each category.

III. Hand Database

In this section we describe the database comprised of hand configurations that we have been collecting as part of the OGR project for LIBRAS. Our first efforts in this sense were to identify which hand configurations are necessary to describe the words in LIBRAS. After a joint study with the Brazilian deaf community, we have arrived at a subset of 103 hand configurations [5], which are depicted in Figure 4.

In order to acquire the 103 symbols, we have developed a mobile application where the hand configuration symbols were presented to the user so that he/she could draw them by copying the template. Up to now about 30 people, deaf and non-deaf, contributed to the database with up to three samples per symbols generating a total of 7,997 grayscale images (480×519) that were stored in PNG (Portable Network Graphics). Figure 5 shows some of the variability of the handwritten samples extracted from the database.

IV. Representations and Classifiers

The similarity with character handwriting recognition problem motivated us to assess state-of-the-art representations and

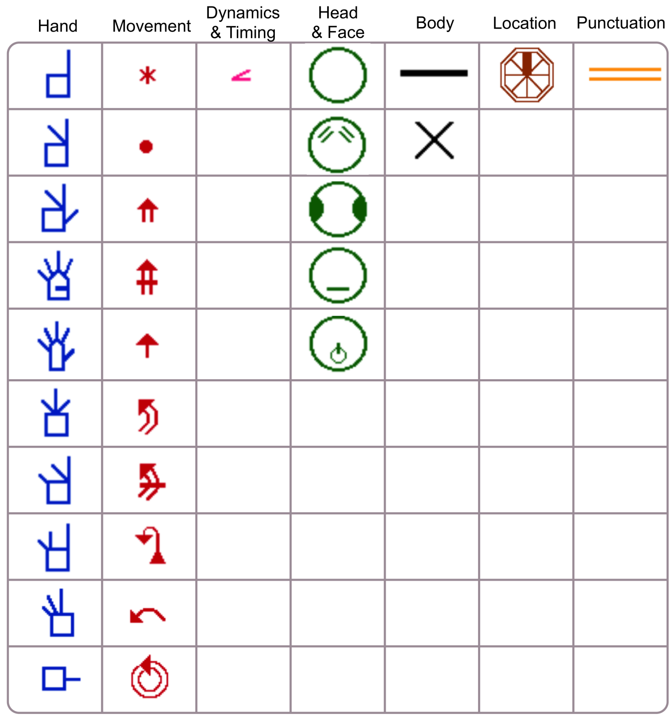


Fig. 3. Example of the 7 categories and 30 groups of the SignWriting.

classifiers used in this field of research. Thus, features such as histogram projections [14], contour profiles [7], and concavity analysis [9] were used to train SVM (Support Vector Machine) classifiers [15].

In addition to the traditional hand-designed feature extraction approach we have also considered a Convolutional Neural Network (CNN) [8], which is a trainable feature extractor and classifier. The literature shows that CNN achieves error rates as low as humans for the task on handwritten digit recognition [2]. This kind of approach is not new, but only recently emerged as a viable alternative due to the appearance and popularization of the Graphic Processing Units (GPUs) which are capable of delivering high computational throughput at relatively low cost, achieved through their massively parallel architecture.

A. Hand-designed Features

Perhaps the simplest feature set used for character recognition is the projection histograms. These features are derived from histograms of horizontal and vertical projections of black pixels. They are extracted from the normalized image of the character so as to obtain normalized histograms of black pixels both on the X-axis as well as on the Y-axis. In our case, the images are cropped and normalised in 32×32 pixels, creating a feature vector of 64 components.

The contour information is extracted from a histogram of contour directions using a zoning mechanism (3 horizontal and 2 vertical zones). For each zone, the contour line segments between neighboring pixels are grouped regarding 8-Freeman directions. The number of line segments of each orientation is counted. Therefore, the contour feature vector is composed of (8×6) 48 components.

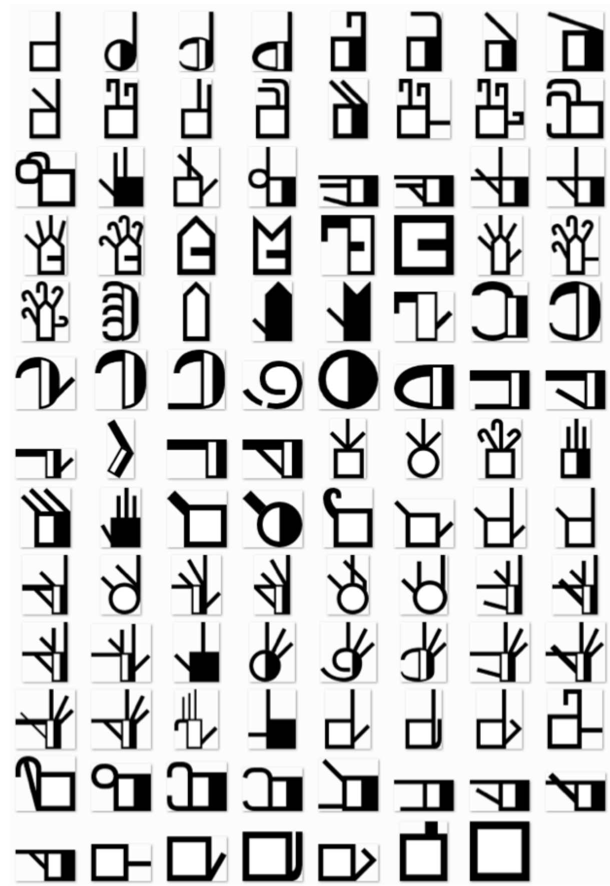


Fig. 4. 103 hand configurations used by the LIBRAS

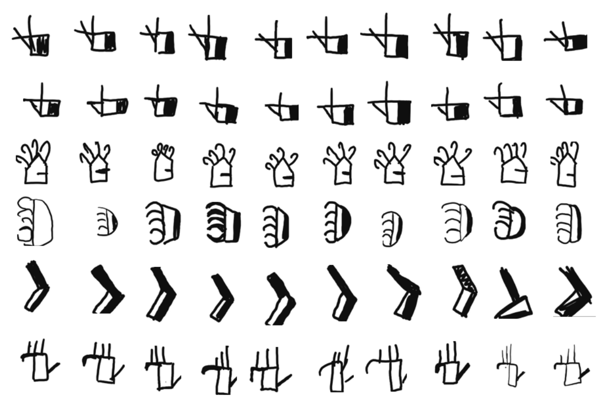


Fig. 5. Handwritten symbols extracted from six classes of the database.

The third hand-designed feature is based on the concavity analysis proposed in [9], which works as follows: for each white pixel in the component, we search in 4-Freeman direction, the number of black pixels that it can reach as well as which directions the black pixel is not reached. When black pixels are reached in all directions, we branch out into four auxiliary directions in order to confirm if the current white pixel is really inside a closed contour. Those pixels that reach just one black pixel are discarded. This generates a feature vector of 13 dimensions. Using the same aforementioned zoning scheme, the concavities are represented by a feature vector of (13×6) 78 components.

B. CNN

The deep neural network architecture used in this research was based on models that achieved high levels of accuracy on object classification tasks. In particular, it contains the repeated use of convolutional layers followed by max-pooling layers, as used by Ciresan et al. [2]. The architecture is illustrated in Figure 6.

In summary, this architecture consists of an input layer (image scaled to 32×32), two combinations of convolutional and pooling layers where each convolutional has 64 filters with size 5×5 and stride set to 1, while the the pooling layers consists of windows with size 3×3 and stride 2. The locally-connected layer has 32 filters of size 3×3 and stride 1. Finally the fully-connected output layer has 103 outputs.

Fully-connected layers are the standard for neural networks, and connect, using unshared weights, all the neurons from one layer to the next one. Locally-connected layers only connect neurons within a small window to the next layer, similarly to convolutional layers, but without sharing weights. Combinations of both types of layers were tested, and the best results were obtained with two locally-connected layers of rectified linear units, and a final fully-connected layer with softmax activation.

V. EXPERIMENTS

The hand-designed features were used to train SVM classifier. Different kernels were tried out, but the best results were achieved using a Gaussian kernel. Parameters C and γ were determined through a grid search. The CNN model was trained on a Tesla C2050 GPU using the cuda-convnet library [2]. Training was stopped when the error on the validation set did not improve in over 100 epochs. All experiments were performed using 3-fold cross validation.

SignWriting has some context, i.e., some hand configurations may not occur together with some head & face and movement configurations, which allows the recognition system the ability to provide a list of symbols that are similar to the queried symbol. The size of this list, also known as the hit list, can vary, e.g. 1, 5, or 10. The results are then expressed in terms of TOP-1, TOP-5, or TOP-10 performance. This means that a hit list will be considered correct if at least one version of the queried symbols appears on it. In this work we report the TOP-1 and TOP-2 performances.

Table II reports the results of the SVM classifiers trained with the hand-designed features. The best recognition rate, 91.6%, was achieved by the classifier trained with the concavity-based features. Table II still reports our efforts in combining the feature vectors and classifiers. First, the feature vectors were combined into a single feature vector that was used to train the SVM classifier. The best combination result, 91.5%, was similar the classifier trained with concavity-based features. Finally, the classifiers were combined through different fusion rules such as Sum, Product, Max, and Voting. These results are in the last part of Table II.

[2] http://code.google.com/p/cuda-convnet/

TABLE II. RESULTS OF THE SVM CLASSIFIERS TRAINED WITH HAND-DESIGNED FEATURES

Feature	TOP-1	TOP-2
(a)Histogram Projections	81.0	91.0
(b)Contour	75.2	87.8
(c)Concavities	91.6	97.4
(d)a+b+c	91.5	97.4
Sum	93.0	98.1
Product	93.9	98.4
Max	91.3	97.4
Voting	92.4	96.2

As one can see, the Sum and Product rules brought some improvement in terms of performance. By analysing the confusion matrices, we have noticed that very few confusions produced by the concavity-based classifier can be solved by other classifiers. This lack of complementarity explains the weak results of the combination rules.

The second part of our experiments were devoted to the CNN classifier. As stated before, one of the advantages of using deep learning techniques is not requiring the design of feature extractors by a domain expert, but instead let the model learn them. We can visualize the feature detectors that model learns on the first convolutional layer, considering the weights on the learned feature maps. Figure 7 displays the 64 feature maps learned on the first convolutional layers of both models.

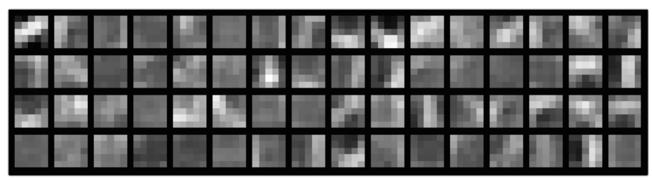


Fig. 7. Feature maps learned by the first convolutional layers.

The results achieved by the CNN classifier was comparable to the results produced by the classifiers trained with the hand-designed features, about 91%. According to Simard et al. [11], if the distribution to be learned has transformation-invariance properties, generating additional data using transformations may improve the performance of the CNN classifier. In the case of handwriting recognition, it has been shown [11] that the distribution has some invariance with respect to not only affine transformations, but also elastic deformations corresponding to uncontrolled oscillations of the hand muscles, dampened by inertial. Therefore, some simple distortions such as translations and rotations can be used to improve the amount of data to train the CNN classifier.

Thus, using such distortions, for each training image we created another 19 images by performing three random crops of size 27×27, three random crops of size 28×28, three rotations to the right (3, 5, and 8 degrees), three rotations to the left (3, 5, and 8 degrees), three smoothed images (mean filter), and two morphological modified images (opening and closing operators with cross structured element). Using this augmented dataset we were able to increase the results in about 3%, as shown in Table III.

As depicted in Figure 4 the large number of classes and the similarity among some classes make this problem an interesting challenge. Figure 8 shows some common confusions to all

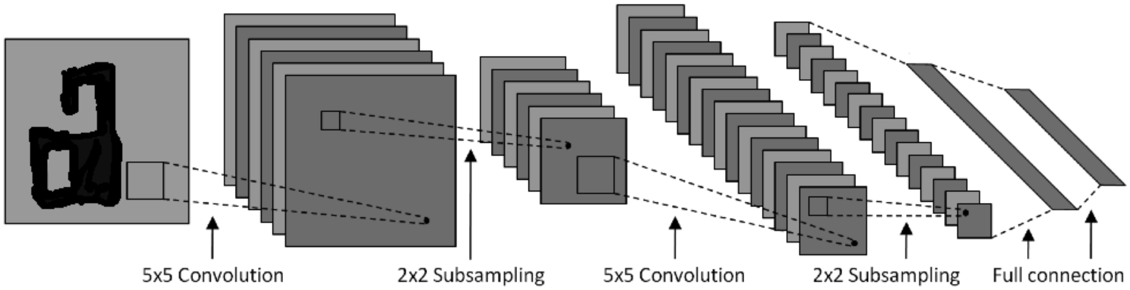
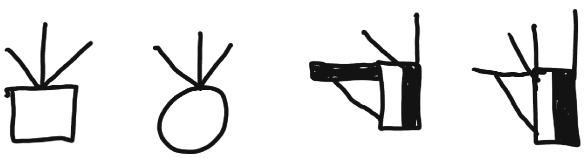


Fig. 6. The Deep Convolutional Network architecture (adapted form [8])

TABLE III. RESULTS OF THE CNN CLASSIFIER

Feature	TOP-1	TOP-2
CNN	91.7	96.5
CNN with data augmentation	94.4	97.8

the classifiers. In both cases, we believe that a strategy based on verifiers specialized to solve specific confusion would help to improve the recognition rates [10].



(a) (b) (c) (d)

Fig. 8. Some confusions common to all classifiers, (a) classified as (b) and (c) classified as (d).

VI. CONCLUSION

In this work we have presented our first efforts towards the automatic classification of SignWriting symbols. As depicted in the examples of Figure 1, automatic classification of SignWriting is quite complex since it involves the classification of hundreds of symbols which may overlap to each other. The results presented on 103 classes of isolated symbols of hand configuration are promising and we believe they can be further improved by using other classifiers, features, and verification schemes. As future work, we intend to continue collecting data so that all the groups (movement, dynamics&timing, head&face, body, location, and punctuation) may be represented in the databases.

ACKNOWLEDGMENT

This research has been supported by The National Council for Scientific and Technological Development (CNPq), grants #151145/2014-8 and #303513/2014-4, and the Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

[1] C. S. Bianchini, F. Borgia, and M. Marsico. Swift - a signwriting editor to bridge between deaf world and e-learning. In *IEEE International Conference on Advanced Learning Technologies*, pages 526–530. IEEE, 2012.

[2] D. Ciresan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3642–3649, 2012.

[3] L. P. Cordella and M. Vento. Symbol recognition in documents: A collection of techniques? *International Journal on Document Analysis and Recognition*, 3(2):73–88, 2000.

[4] I. Fajardo, M. Vigo, and L. Salmeron. Technology for supporting web information search and learning in sign language. *Interacting with Computers*, 21(4):243–256, 2009.

[5] C. Guimarães, J. F. Guardezi, and S. Fernandes. Sign language writing acquisition - technology for a writing system. In *47th Hawaii International Conference on System Science*, pages 120–129. IEEE, 2014.

[6] C. Guimarães and S. Fernandes J. F. Guardezi, L. S. Oliveira. Deaf culture and sign language writing system: a database for a new approach to writing system recognition technology. In *47th Hawaii International Conference on System Science*, pages 3368–3377, 2014.

[7] L. Heutte, T. Paquet, J. V. Moreau, Y. Lecourtier, and C. Olivier. A structural/statistical feature based vector for handwritten character recognition. *Pattern Recognition Letters*, 19:629–641, 1998.

[8] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Procs of IEEE*, 86:2278–2324, 1998.

[9] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen. Automatic recognition of handwritten numerical strings: A recognition and verification strategy. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(11):1438, 1454 2002.

[10] L. S. Oliveira, R. Sabourin, F. Bortolozzi, and C. Y. Suen. Impacts of verification on a numeral string recognition system. *Pattern Recognition Letters*, 24(7):1023–1031, 2003.

[11] P. Y. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *International Conference on Document Analysis and Recognition*, pages 958–963, 2003.

[12] V. Sutton. A way to analyze american sign language and any other sign language without translation into any spoken language. In *National Symposium on Sign Language Research and Teaching*, 1980.

[13] K. Tombre, S. Tabbone, and P. Dosch. Musings on symbol recognition. In *6th International Conference on Graphics Recognition*, pages 23–34, 2006.

[14] O. D. Trier, A. K. Jain, and T. Taxt. Feature extraction methods for character recognition - a survey. *Pattern Recognition*, 29(4):641–662, 1996.

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