Handwritten Gujarati Digit Recognition using Sparse Representation Classifier

Kamal Moro, Mohammed Fakir and Belaid Bouikhalene

Department of Computers Sciences Sultan Moulay Slimane univefsity Beni Mellal, Morocco

Abstract:*We present in this paper a framework for handwritten Gujarati digit recognition using Sparse Representation Classifier. The classifier assumes that a test sample can be represented as a linear combination of the train samples from its native class. Hence, a test sample can be represented using a dictionary constructed from the train samples. The most sparse linear representation of the test sample in terms of this dictionary can be efficiently computed through l_1-minimization, and can be exploited to classify the test sample. This is a novel approach for Gujarati Optical Character Recognition, and demonstrates an accuracy of 80,33%. This result is promising, and should be investigated further.*

Keywords:*Sparse Representation Classifier, Gujarati Optical Character Recognition, Handwritten character recognition, Digit recognition.*

1. Introduction

Optical character recognition (OCR) is verypopular research field since 1950's, and a large number of researches has been devoted to this subject. Character recognition systems offer potential advantages by providing an interface that facilitates interaction between man and machine, these systems are based on algorithms consisting essentially of three essential phases: pretreatment, feature extraction and classification.

The classification of digital documents is a complex task in a document analysis flow. The number of documents resulting from the OCR retro-conversion (optical character recognition) makes the classification task harder. In the literature, different features are used to improve the classification quality; one of the most recent is sparse representation. This method received a great deal of attentions in recent years.One basic idea in compressed sensing is that most signals have a sparse representation as a linear combination of a reduced subset of signals from the same space. Naturally the signals tend to have a representation biased towards their own class, i.e. the sparse representation is mainly formed from samples from its own class. This is the starting point for Sparse Representation based Classification (SRC)[1], which proved to be a state-of-the art method for face recognition.

This paper proposes the use of Sparse Representation Classifier (SRC), and has shown to achieve an accuracy of 80,33% as the aforementioned classifier is applied on a data set of 600 samples. SRC is first used in face recognition problem [1], and since then, has been applied in various classification problems such as vehicle tracking [2], speaker recognition [3], classification of hyperspectral imagery [4], and cancer detection [5]. To the best of our knowledge, SRC has never been applied to the problem of Gujarati handwritten digit recognition.

2. Sparse Representation based Classification

2.1. Principe

A basic problem in object recognition is to use labeled training samples from k distinct object classes to correctly determine the class to which a new test sample belongs. We arrange the given n_i training samples from the i^{th} class as columns of a matrix$A_i = [v_{i1}, v_{i2}, \dots, v_{in_i}]$.

Given this sufficient training samples of the ith object class, any new (test) sample $y \in R^m$ from the same class will approximately lie in the linear span of the training samples associated with object i:

$$y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \cdots + \alpha_{i,n_i}v_{i,n_i} \quad (1)$$

for some scalars, $\alpha_{i,j} \in R, j = 1,2, \dots, n_i$.

So, we define a new matrix A for the entire training set as the concatenation of the n training samples of all k object classes:

$$A = [A_1, A_2, \dots, A_k] = [v_{1,1}, v_{1,2}, \dots, v_{k,n_k}] \quad (2)$$

Then the linear representation of y can be rewritten in the terms of all training samples as

$$y = Ax_0 \in R^m \quad (3)$$

where $x_0 = [0, \dots, 0, \alpha_{i,1}, \alpha_{i,2}, \dots, \alpha_{i,n_i}, 0, \dots, 0]^T \in R^m$ is a coefficient vector whose entries are zero except those associated with the i[th] class.

As the entries of the vector x_0 encode the identity of the test sample y, it is tempting to attempt to obtain it by solving the linear system of equations $y = Ax$. Obviously, if $m > n$, the system of equations $y = Ax$ is overdetermined, and the correct x_0 can usually be found as its unique solution, however, in the OCR, the system $y = Ax$ is typically underdetermined, and so, its solution is not unique. Conventionally, this difficulty is resolved by choosing the minimum l^2-norm solution:

$$\widehat{x_2} = \arg\min ||x||_2 \quad \text{subject to} \quad Ax = y \quad (4)$$

While this optimization problem can be easily solved (via the pseudo inverse of *A*), the solution $\hat{x}_2$ is generally is not especially informative for recognizing the test sample *y*. This motivates us to seek the sparsest solution to $y = Ax$, solving the following optimization problem:

$$\widehat{x_0} = \arg\min ||x||_0 \quad \text{subject to} \quad Ax = y \quad (5)$$

where $||.||_0$ denotes the l^0-norm, which counts the number of nonzero entries in a vector. However, the problem of finding the sparsest solution of an underdetermined system of linear equations is hard and difficult even to approximate [6], so, some development in the emerging theory of sparse representation and compressed sensing ([7], [8] [9]) reveals that if the solution x_0 sought is sparse enought, the solution of the l^0-minimization problem (5) is equal to the solution to the following l^1-minimization problem:

$$\widehat{x_1} = \arg\min ||x||_1 \quad \text{subject to} \quad Ax = y \quad (6)$$

2.2. Dealing with small dense noise

So far, we have assumed that (3) holds exactly. Since real data are noisy, it may not be possible to express the test sample exactly as a sparse superposition of the training samples. The model (3) can be modified to explicitly account for small possibly dense noise by writing

$$y = Ax_0 + z \quad (7)$$

where $z \in R^m$ is a noise term with $||z||_2 < \varepsilon$. The sparse solution x_0 can still be approximately recovered by solving the following stable l^1-minimization problem:

$$\widehat{x_1} = \arg\min ||x||_1 subject to ||Ax - y||_2 < \varepsilon \quad (8)$$

2.2. Algorithm

For each class i, let $\delta_i : R^n \to R^n$ be the characteristic function that selects the coefficients associated with the*i*[th] class. For $x \in R^n, \delta_i(x) \in R^n$ is a new vector whose only nonzero entries are the entries in *x* that are associated with the i^{th} class, one can approximate the given test sample *y* as $\widehat{y_i} = A\delta_i(\widehat{x_1})$. We then classify y based on these approximations by assigning it to the object class that minimizes the residual between y and $\widehat{y_i}$:

$$\min_i r_i(y) = ||y - A\delta_i(\widehat{x_1})||_2 \quad (9)$$

Algorithm.

1. **Input:** a matrix of training samples $A = [A_1, A_2, \dots, A_k] \in R^{m*n}$ for k classes, a test sample $y \in R^m$, (and an optional error tolerance $\varepsilon > 0$).
2. Solve the l^1-minimization problem: $\widehat{x_1} = argmin_x ||x||_1$ subject to $Ax = y$ (5) (or alternatively, solve $\widehat{x_1} = argmin_x ||x||_1$ subject to $||Ax - y||_2 < \varepsilon$).
3. Compute the residuals $r_i(y) = ||y - A\delta ix1/2$ for *i=1,…,k.*
4. **Output:** $identify(y) = argmin_i r_i(y)$.

3. Gujarati database

Gujarati belonging to Devnagari family of languages, which originated and flourished in Gujarat a western state of India, is spoken by over 50 million people of the state. Though it has inherited rich cultural and literature, and is a very widely spoken language, hardly any significant work has been done for the identification of Gujarati optical characters. The Gujarati script differs from those of many other Indian languages not having any shirolekha (headlines). Gujarati numerals do not carry shirolekha and it applies to almost all Indian languages. The numerals in Indian languages are based on sharp curves and hardly any straight lines are used. Figure.1 is a set of Gujarati numerals.

Gujarati digits (Figure 1) are very peculiar by nature. Only two Gujarati digits one(1) and five(5) are having straight line, making Gujarati digit identification a little more difficult. Also Gujarati digits often invite misclassification. These confusing sets of digits' areas shown in Figure 2.

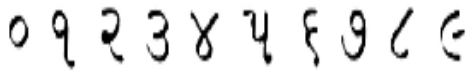


Figure 1. Gujarati digits 0-9

Figure 2. Confusing Gujarati digits

For developing a system to identify Gujarati handwritten digits, we have collected numerals 0-9 written in Gujarati scripts from a large number of writers. These numbers were scanned in 300 dpi by a flatbed scanner. Initially they are in separate boxes of 50*30 pixels each. Since our problem is to identify handwritten digits, the first thing required is to bring all the characters in a standard normal form. This is needed because when a writer writes he may use different types of pens, papers, they may follow even different styles of writing etc.

4. Experimental Results

For this experiment, this recognition process was trained for total 30 sets of digits, and was tested for 60 other new sets of digits. In total the network was trained by 300 digits and tested for 600 digits. Figure.3 shows the recognition process used.

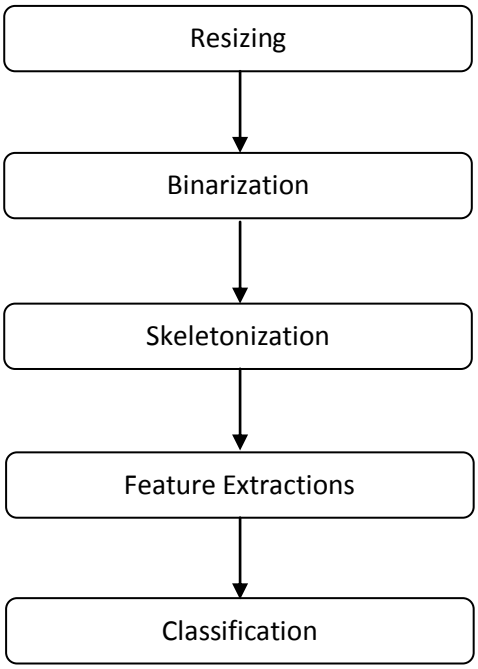
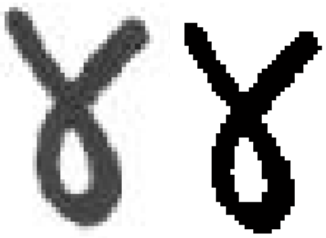


Figure 3. Recognition Process

In terms of classification, we performed recognition on 600 characters, so 60 characters for each class, based on 300 learning characters so 30 characters for each class.

4.1. Binarization

Binarization is often the first step in treatment systems and image analysis ([10], [11], [12]) especially images of documents. Its goal is to reduce the amount of information present in the image, and keeping only the relevant information, which allows us to use simple analysis methods vis-à-vis images in grayscale or color [13]. In our case, we used the algorithm of Wolf [14] (Figure.4).



a *b*

Figure.4. Binarization of a digit, a: before binarization, b: after binarization

4.2. Skeletonization

In the continuous scheme, a skeleton of the form is a plurality of lines passing down the middle (Figure.5). The objective of skeletonization is to represent a set with a minimum of information in a form that is as simple to remove and convenient to handle. This is the notion of a continuous centerline form introduced by Blum [15] who defined a medial axis for computing a skeleton of a shape, using an intuitive model of fire propagation on a grass field, where the field has the form of the given shape. If one "sets fire" at all points on the boundary of that grass field simultaneously, then the skeleton is the set of quench points, i.e., those points where two or more wave fronts meet. This intuitive description is the starting point for a number of more precise definitions.

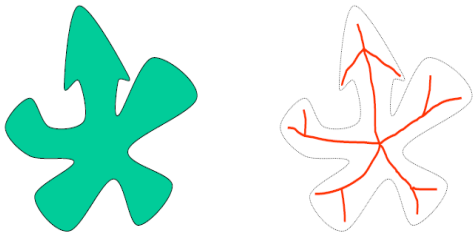
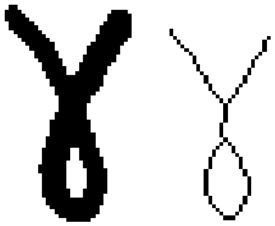


Figure 5. Example of Skeleton

There are currently a variety of methods to build skeletons from forms. One of the best known is the topological thinning, it is to examine the pixels of the binary image and iteratively remove those who do not belong to the final skeleton. Reviews of pixels is done in two ways: sequential approach where the removal of a point P to the *nth* iteration depends at transactions made so far but also the pixels already processed and parallel approach in which the pixels are examined independently at each iteration, the removal of the P to the *nth* iteration depends at operations performed at the previous iteration. It is the latter approach which is used by applying the algorithm Guo_Hall [16] (Figure.6). This methodisused in [17] and [18].



a *b*

Figure 6. Skeleton of a digit, a: beforeskeletonizing, b: after skeletonizing

4.3. Feature Extraction

We use the Box-approach in Refs [17], [18], [19], [20]. This approach requires the special division of the character image. The major advantage of this approach stems from its robustness to small variations, ease of implementation and relatively high recognition rate. The choice of box size and number of boxes is discussed in section 6 on results.

Each character image is divided into 24 boxes so that the portions of a numeral will be in some of these boxes. There could be boxes that are empty, as shown in Figure.7. English numeral 3 is enclosed in the 6*4 grid. However, all boxes are considered for analysis in a sequential order. By considering the bottom left corner as the absolute origin (0,0), the coordinate distance (vector distance) for the *k*th pixel in the b^{th} box at location (i,j) is computed as (10)

$$d_{kb} = \sqrt{(i^2 + j^2)} \qquad (10)$$

By dividing the sum of distances of all black pixels present in a box with the total number of pixels in that box, a normalized vector distance for each box is obtained as

$$\gamma_b = \frac{1}{n_b} \sum_{k=1}^{n_b} d_k^b, \qquad b = 1,2, \dots 24 \qquad (11)$$

Where n_b is the total number of pixels in b^{th}box. These vector distances constitute a set of features based on distances. Therefore, $24\gamma_b$'s corresponding to 24 boxes will constitute a feature set. However, for empty boxes, the value will be zero.

In practice, each character is divided into 50 blocks of 10*5 pixels. A sample of the extraction vectoris as follows:

[0 0 0.6672 0.7017 0 0 1.9429 2.8865 0 0 0.5770 2.0683 0 0 0 3.3593 0 0 1.1157 0 2.4256 1.8303 0 2.4254 3.5686 0 4.0410 6.6789 9.9418 0 0 0 0 7.1634 0 0 0 0 8.0269 0 0 9.6223 0 8.6789 0 0 0 10.4075 3.5689 0].

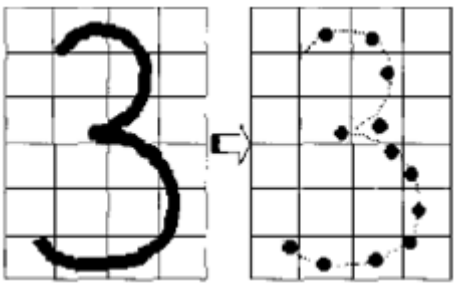


Figure 7. Portions of the numeral lie within some boxes while others are empty

4.4. Results

Following the recognition process of Figure.2, 482 characters are recognized among 600 characters, so a recognition rate of 80.33%, the Table.1 gives more details on the recognition rate for each class.

We notice, in general, most characters are poorly confused one another, except for characters '6' and'3', and especially for characters '8' and '9', due the large similarity of this digits.

Figures.8 shows the evolution of the recognition rate by varying the number of training characters.

According to this scheme, we see that the number of testing character recognized is proportional by increasing the number of training character.

Table 1. Performance of the process

Numbers	0	1	2	3	4	5	6	7	8	9	Success (%)
0	50	0	0	0	0	0	0	1	7	2	83,33
1	0	42	5	0	1	10	1	0	1	0	70
2	1	4	52	1	0	0	0	0	2	0	86,67
3	0	0	0	59	0	0	0	1	0	0	98,33
4	0	0	1	1	50	0	0	1	7	0	83,33
5	0	1	7	0	1	50	1	0	0	0	83,33
6	0	1	0	16	0	1	41	0	1	0	68,33
7	2	0	0	7	0	0	1	47	3	0	78,33
8	0	0	0	0	0	0	0	0	60	0	100
9	1	0	0	1	0	0	0	0	27	31	51,67

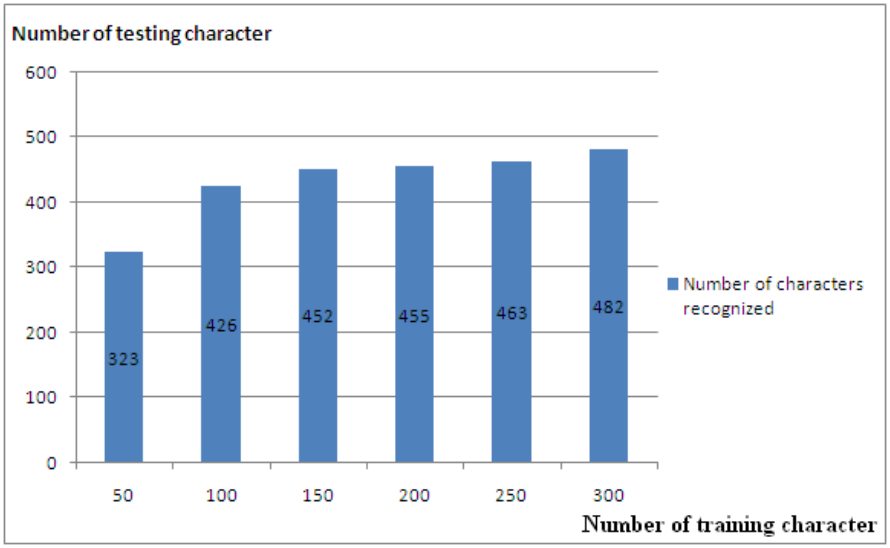


Figure.8. Evolution of the recognition rate by varying the number of training characters

5. Conclusion

We used a novel classifier, Sparse Representation Classifier (SRC), for off-line isolated handwritten Gujarati numeral recognition. The method demonstrates a good result with 80,33% overall accuracy. This result is very promising, and is likely to improve if another feature extraction techniqueis applied. Comparison with other conventional classifiers should be considered in future as a continuation of this work. The results should also be verified for other standard handwritten character databases such as the ISI database of handwritten Bangala numerals.

References

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, 2009.

[2] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 33, no. 11, pp. 2259–2272, 2011.

[3] J. M. K. Kua, E. Ambikairajah, J. Epps, and R. Togneri, "Speaker verification using sparse representation classification," in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on. IEEE, 2011, pp. 4548–4551.

[4] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 10 PART 2, pp. 3973–3985, 2011.

[5] A. A. Helal, K. I. Ahmed, M. S. Rahman, and S. K. Alam, "Breast cancer classification from ultrasonic images based on sparse representation by exploiting redundancy," in 16th International Conference on Computer and Information Technology, ICCIT 2013, Mar. 2014.

[6] E. Amaldi and V. Kann, "On the Approximability of Minimizing Nonzero Variables or Unsatisfied Relations in Linear Systems," Theoretical Computer Science, vol. 209, pp. 237-260, 1998.

[7] D. Donoho, "For Most Large Underdetermined Systems of Linear Equations the Minimal l1-Norm Solution Is Also the Sparsest Solution," Comm. Pure and Applied Math., vol. 59, no. 6, pp. 797- 829, 2006.

[8] E. Cande`s, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," Comm. Pure and Applied Math., vol. 59, no. 8, pp. 1207-1223, 2006.

[9] E. Cande`s and T. Tao, "Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?" IEEE Trans. Information Theory, vol. 52, no. 12, pp. 5406-5425, 2006.

[10] Trier.Q. D and Taxt.T. Evaluation of binarization methods for document images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 17(3), pp. 312-315, 1995

[11] Leedham.G ,Varma.S , Patankar.A, and Govindaraju.V. Separating text and background in degraded documents images - A comparison of global thresholding techniques for multi-stage thresholding. Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition, pp. 244–249, 2002

[12] Kefali.A, Sari.T and Sellami.M. Evaluation de plusieurs Techniques de seuillage d'images de documents arabes anciens. IMAGE'09 Biskra, 2009

[13] Sauvola.J, Pietikainen.M. Adaptive document image binarization. Pattern Recognition, 33(2), pp. 225–236, 2000

[14] Wolf.C, Jolion.J.M, and Chassaing.F. Extraction de texte dans des vidéos: le cas de la binarisation, 13ème Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle, pp. 145-152, 2002

[15] Blum.H. A transformation for extracting new descriptions of shape.Models for the Perception of Speech and Visual Form, MIT Press, pp. 362–380, 1967

[16] Guo.Z and Hall.R.W. Parallel thinning with two-subiteration algorithms. *Comm. ACM*, 32(3) :359–373, 1989

[17] M. Hanmandlu.M, O.V. Ramana.M.O.V . Fuzzy model based recognition of handwritten numerals. Pattern Recognition 40, pp: 1840-1852, 2007

[18] Moro.K, Fakir.M, El Kessab.B, Bouikhalene.B, Daoui.C. Comparison of two feature extraction methods based on the raw form and his skeleton for gujarati handwritten digits. FACTA UNIVERSITATIS (NIˇS) Ser. Math. Inform. Vol. 28, No 2, 161–178, 2013

[19] Hanmandlu.M, Murali Mohan.K.R, Chakraborty.S, Goyal.S, Roy Choudhury.D. (2003). Unconstrained handwritten character recognition based on fuzzy logic. Pattern Recognition 36 (3), pp: 603-623.

[20] Hanmandlu.M, Yusof.M.H.M, VamsiKrishna.M. (2005). Off line signature verification and forgery detection using fuzzy modeling. Pattern Recognition 38 (3), pp: 341-356.

Kamal Moro obtained a MS degree and a PHD degree in Computer Sciences from Sultan Moulay Slimane University in 2009 and 2016 respectively. He is currently with the national institute of commerce and organization, Settat, Morocco. His research interest includes character recognition and image processing.

Belaid Bouikhaleneobtained a Ph.D. degree in Mathematics in 2001 and a degree of Master in Telecommunication and Computer science in 2007 from the University of Ibn TofelKenitra, Morocco. He is currently a professor at University Sultan MoulaySlimane, Morocco, His research focuses on mathematics and applications, decision information systems, e-learnig, pattern recognition and artificial intelligence.

Mohamed Fakirreceived the B.Sc. degree in physics (option Electronics) from Mohamed V University, Rabat, Morocco in (1984), Master Degree from Nagoka University of Technology in 1991 and PHD degree from Cadi AyyadUniversity, Morocco, in 2001.He worked with Hitachi Ltd, Japan from 1991 to 1994 in air conditioning design department. He is currently a professor at University Sultan MoulaySlimane, Morocco, His research includes image processing, datamining, web mining and pattern recognition