ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

Text-Attentional Convolutional Neural Network for Scene Text Detection

Ankita Wankhade, Kalpana Malpe

M. E Student, Department of Computer Science and Engineering, Gurunanak Institute of Engineering and Technology, Nagpur, India

Professor, Department of Computer Science and Engineering, Gurunanak Institute of Engineering and Technology, Nagpur, India

ABSTRACT: This process presents research aimed at achieving better OCR quality in large scale digitization of newspapers and books, and opening possibilities of full-text search of digitized old Slovenian printed texts, which should enable digital library end-users to gain better transcriptions of digitized contents. Segmentation of handwritten document images into text-lines and words is an essential task for optical character recognition. However, since the features of handwritten document are irregular and diverse depending on the person, it is considered a challenging problem. In order to address the problem, we formulate the word segmentation problem as a binary quadratic assignment problem that considers pairwise correlations between the gaps as well as the likelihoods of individual gaps. Even though many parameters are involved in our formulation, we estimate all parameters based on the Structured CNN framework so that the proposed method works well regardless of writing styles and written languages without user-defined parameters. Experimental results on ICDAR 2009/2013 handwriting segmentation databases show that proposed method achieves the state-of-the-art performance on Latin-based and Indian languages. But in this paper there is the problem of word spotting and word recognition on images. In the proposed system, this is achieved by a combination of label embedding and attributes learning, and a common subspace regression. Then the images and strings represent the same word which are close to each other allowing one to cast recognition and retrieval tasks. Compared with the existing method, the advantage of our method has a fixed length, low dimensional and very fast to compute. In the preprocessing the given dataset is filtered by using median filter. After that, in the segmentation process every image is cropped identically by the bounding box segmentation. Then in the feature extraction is done by Gabor wavelet for each and every character which is cropped from bounding box. For classifying the image we use SVM classifier.Matlab software and its image processing toolbox have been used in images processing and analysis. We test our approach for the given dataset of both handwritten documents and natural images showing results comparable or better than the state-of-the-art on spotting and recognition tasks.

KEYWORDS: Pre-Processing, Feature Extraction, Segmentation, Classification, Recognition, SVM

I. INTRODUCTION

In this process a complete OCR methodology for recognizing historical documents, either printed or handwritten without any knowledge of the font, is presented. This methodology consists of three steps: The first two steps refer to creating a database for training using a set of documents, while the third one refers to recognition of new document images. First, a pre-processing step that includes image Binarization and enhancement takes place. At a second step a top-down segmentation approach is used in order to detect text lines, words and characters. A clustering scheme is then adopted in order to group characters of similar shape. This is a semi-automatic procedure since the user is able to interact at any time in order to correct possible errors of clustering and assign an ASCII label. After this step, a database is created in order to be used for recognition. Finally, in the third step, for every new document image the above segmentation approach takes place while the recognition is based on the character database that has been produced at the previous step.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

TEXT understanding in images is an important problem that has drawn a lot of attention from the computer vision community since its beginnings. Text understanding covers many applications and tasks, most of which originated decades ago due to the digitalization of large collections of documents. This made necessary the development of methods able to extract information from these document images: layout analysis, information flow, transcription and localization of words, etc. Recently, and motivated by the exponential increase of publicly available image databases and personal collections of pictures, this interest now also embraces text understanding on natural images.
Methods able to retrieve images containing a given word or to recognize words in a picture have also become feasible and useful. In this paper we consider two problems related to text understanding: word spotting and word recognition. In word spotting, the goal is to find all instances of a query word in a dataset of images. The query word may be a text string – in which case it is usually referred to as query by string (QBS) or query by text (QBT) –, or may also be an image, – in which case it is usually referred to as query by example (QBE). In word recognition, the goal is to obtain a transcription of the query word image. In many cases, including this work, it is assumed that a text dictionary or lexicon is supplied at test time, and that only words from that lexicon can be used as candidate transcriptions in the recognition task. In this work we will also assume that the location of the words in the images is provided, i.e. we have access to images of cropped words. If those were not available, text localization and segmentation techniques could be used, but we consider that out of the scope of this work.Due to the great amount of variability in handwriting and the high noise levels in historical documents, handwritten historical documents are currently transcribed by hand. In essence, this means that each occurrence of a word in a corpus must be annotated. The goal of the Word Spotting idea applied to handwritten documents is to greatly reduce the amount of annotation work that has to be performed, by grouping all words into clusters. Once such a clustering of the data set exists, the number of words contained in a cluster can be used as a cue for determining the importance of the word as a query term. For example, highly frequent terms, such as the, of, etc. are stop words and can be discarded.

II. LITERATURE SURVEY

1) Title: Handwritten Word Spotting with Corrected Attributes (2013)
Author: J. Almazan, A. Gordo, A. Fornes

We propose an approach to multi-writer word spotting, where the goal is to find a query word in a dataset comprised of document images. We propose an attributes-based approach that leads to a low-dimensional, fixed-length representation of the word images that is fast to compute and, especially, fast to compare. This approach naturally leads to an unified representation of word images and strings, which seamlessly allows one to indistinctly perform queryby-example, where the query is an image, and query-bystring, where the query is a string. We also propose a calibration scheme to correct the attributes scores based on Canonical Correlation Analysis that greatly improves the results on a challenging dataset. We test our approach on two public datasets showing state-of-the-art results.

ADVANTAGE:

- We believe that sharing information between words is extremely important to learn good discriminative representations, and that the use of attributes is one way to achieve this goal.

DISADVANTAGE:

- Where the query is an image of a handwritten word and it is assumed that the transcription of the dataset and the query word is not available.

2. Title: A scale space approach for automatically segmenting words from historical handwritten documents (2005)
Author: R. Manmatha and J. Rothfeder

Many libraries, museums, and other organizations contain large collections of handwritten historical documents, for example, the papers of early presidents like George Washington at the Library of Congress. The first step in providing recognition/retrieval tools is to automatically segment handwritten pages into words. State of the art segmentation techniques like the gap metrics algorithm have been mostly developed and tested on highly constrained

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

documents like bank checks and postal addresses. There has been little work on full handwritten pages and this work has usually involved testing on clean artificial documents created for the purpose of research. Here, a novel scale space algorithm for automatically segmenting handwritten (historical) documents into words is described. First, the page is cleaned to remove margins. This is followed by a gray-level projection profile algorithm for finding lines in images. Each line image is then filtered with an anisotropic Laplacian at several scales. This procedure produces blobs which correspond to portions of characters at small scales and to words at larger scales. Crucial to the algorithm is scale selection that is, finding the optimum scale at which blobs correspond to words.

ADVANTAGE:

- A post processing filtering step is performed to eliminate boxes of unusual size which are unlikely to correspond to words.

DISADVANTAGE:

- Historical manuscript images, on the other hand, contain a great deal of noise and are much more challenging.

3. Title: Real-time scene text localization and recognition (2012)

Author: L. Neumann and J. Matas

An end-to-end real-time scene text localization and recognition method is presented. The real-time performance is achieved by posing the character detection problem as an efficient sequential selection from the set of Extremal Regions (ERs). The ER detector is robust to blur, illumination, color and texture variation and handles lowcontrast text. In the first classification stage, the probability of each ER being a character is estimated using novel features .Only ERs with locally maximal probability are selected for the second stage, where the classification is improved using more computationally expensive features. A highly efficient exhaustive search with feedback loops is then applied to group ERs into words and to select the most probable character segmentation. Finally, text is recognized in an OCR stage trained using synthetic fonts. The method was evaluated on two public datasets. On the ICDAR 2011 dataset, the method achieves state-of-the art text localization results amongst published methods and it is the first one to report results for end-to-end text recognition. On the more challenging Street View Text dataset, the method achieves state-of-the-art recall. The robustness of the proposed method against noise and low contrast of characters is demonstrated by “false positives” caused by detected watermark text in the dataset.

ADVANTAGE:

- The connected component methods is that their complexity typically does not depend on the properties of the text.

DISADVANTAGE:

- Their disadvantage is a sensitivity to clutter and occlusions that change connected component structure.

4. Title: Scene text localization and recognition with oriented stroke detection (2013)

Author: L. Neumann and J. Matas

An unconstrained end-to-end text localization and recognition method is presented. The method introduces a novel approach for character detection and recognition which combines the advantages of sliding-window and connected component methods. Characters are detected and recognized as image regions which contain strokes of specific orientations in a specific relative position, where the strokes are efficiently detected by convolving the image gradient field with a set of oriented bar filters. Additionally, a novel character representation efficiently calculated from the values obtained in the stroke detection phase is introduced. The representation is robust to shift at the stroke level, which makes it less sensitive to intra-class variations and the noise induced by normalizing character size and positioning.

ADVANTAGE:

- The detected strokes induce the set of rectangles to be classified, which reduces the number of rectangles by three orders of magnitude when compared to the standard sliding-window methods.

DISADVANTAGE:

- Such methods is a dependence on the assumption that a character is a connected component, which is very brittle - a change in a single image pixel introduced by noise can cause an unproportional change in the connected component size, shape or other properties, which subsequently affects its classification.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

5. Title: A novel word spotting method based on recurrent neural networks (2012)
Author: V. Frinken, A. Fischer, R. Manmatha

Keyword spotting refers to the process of retrieving all instances of a given keyword from a document. In the present paper, a novel keyword spotting method for handwritten documents is described. It is derived from a neural network-based system for unconstrained handwriting recognition. As such it performs template-free spotting, i.e., it is not necessary for a keyword to appear in the training set. The keyword spotting is done using a modification of the CTC Token Passing algorithm in conjunction with a recurrent neural network. We demonstrate that the proposed systems outperform not only a classical dynamic time warping-based approach but also a modern keyword spotting system, based on hidden Markov models. Furthermore, we analyze the performance of the underlying neural networks when using them in a recognition task followed by keyword spotting on the produced transcription. We point out the advantages of keyword spotting when compared to classic text line recognition.

ADVANTAGE:

- The most effective use of RNNs for sequence labelling is to combine them with HMMs in the so-called hybrid approach.

DISADVANTAGE:

- The problem is that the standard neural network objective functions are defined separately for each point in the training sequence.

III. EXISTING SYSTEM

This work presents an Offline Cursive Word Recognition System dealing with single writer samples. The system is based on a continuous density Hidden Markov Model trained using either the raw data, or data transformed using Principal Component Analysis or Independent Component Analysis. Both techniques significantly improved the recognition rate of the system. Preprocessing, normalization and feature extraction are described as well as the training technique adopted.

Several experiments were performed using a publicly available database. The accuracy obtained is the highest presented in the literature over the same data. The system is based on a sliding window approach: a window shifts column by column across the image and, at each step, isolates a frame. A feature vector is extracted from each frame and the sequence of frames so obtained is modeled with Continuous Density Hidden Markov Models (HMMs). The use of the sliding window approach has the important advantage of avoiding the need of an independent segmentation, a difficult and error prone process.

In order to reduce the number of parameters in the HMMs, we use diagonal covariance matrices in the emission probabilities. This corresponds to the unrealistic assumption of having de correlated feature vectors. For this reason, we applied Principal Component Analysis (PCA) and Independent Component Analysis (ICA) to de-correlate the data. This allowed a significant improvement of the recognition rate. The recognition accuracy achieved with the approach proposed here is, to our knowledge, the highest among the results over the same data presented in the literature.

The analysis of the recognition as a function of the word length shows that the system achieves a recognition rate for samples longer than six letters. This suggests that the performance of our system in tasks involving words with high average length can be very good. Both PCA and ICA had a positive effect on the recognition rate, PCA in particular reduced the error rate. A further improvement can probably be obtained by using nonlinear or kernel PCA. Such techniques often work better than the linear transform we used to perform PCA.

The use of data dependent heuristics was avoided in order to make the system flexible with respect to a change of writer. Any ad-hoc algorithm for the specific style of the writer was avoided. The prior information about the word frequency and distribution can be useful to improve the recognition of short words. These are typically articles, conjunctions and propositions that appear often in the sentences. For this reason, a possible future direction to follow is the application of language models that take into account this kind of information.

DISADVANTAGE:

- In order to reduce the number of parameters in the HMMs, we use diagonal covariance matrices in the emission probabilities. This corresponds to the unrealistic assumption of having de correlated feature vectors.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

- It is necessary to segment the words into letters to perform the training and information loss may occurs.

IV. PROPOSED SYSTEM

In the proposed system, this is achieved by a combination of label embedding and attributes learning, and a common subspace regression. Then the images and strings represent the same word which are close to each other allowing one to cast recognition and retrieval tasks. Compared with the existing method, the advantage of our method has a fixed length, low dimensional and very fast to compute.Word spotting in document images has attracted attention in the document analysis and still poses lots of challenges due to the difficulties of historical documents, different scripts, noise, handwritten documents, etc.

Regarding word recognition, handwritten recognition still poses an important challenge for the same reasons. A model is first trained using labeled training data. At test time, given an image word and a text word, the model computes the probability of that text word being produced by the model when fed with the image word. Recognition can then be addressed by computing the probabilities of all the lexicon words given the query image and retrieving the nearest neighbor. As in the word spotting case, the main drawback here is the comparison speed, since computing these probabilities is orders of magnitude slower than computing a Euclidean distance or a dot product between vectorial representations.

The increasing interest in extracting textual informationfrom real scenes is related to the recent growth of image databases such as Google Images or Flickr. Some interesting tasks have been recently proposed, e.g., localization and recognition of text in Google Street View images or recognition in signs harvested from Google Images. The high complexity of these images when compared to documents, mainly due the large appearance variability, makes it very difficult to apply traditional techniques of the document analysis field. However, with the recent development of powerful computer vision techniques some new approaches have been proposed.To learn how to retrieve and recognize words that have not been seen during training, it is necessary to be able to transfer knowledge between the training and testing samples. One of the most popular approaches to perform this zero shot learning in computer vision involves the use of visual attributes in our case, we propose a broader framework since we do not constrain the choice of features or the method to learn the attributes.

ADVANTAGE

- Our learning approach is currently based on whole word images and does not require to segment the individual characters of the words during training or test, and leads to the large improvements in accuracy.
- This method does not require an available lexicon for a full recognition of the image words.

V. ALGORITHM EXPLANATION

Networks (CNNs / ConvNets)

Convolutional Neural Networks are very similar to ordinary Neural Networks from the previous chapter: they are made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product and optionally follows it with a non-linearity. The whole network still expresses a single differentiable score function: from the raw image pixels on one end to class scores at the other. And they still have a loss function (e.g. SVM/Softmax) on the last (fully-connected) layer and all the tips/tricks we developed for learning regular Neural Networks still apply.So what does change? ConvNet architectures make the explicit assumption that the inputs are images, which allows us to encode certain properties into the architecture. These then make the forward function more efficient to implement and vastly reduce the amount of parameters in the network.

In this way, ConvNets transform the original image layer by layer from the original pixel values to the final class scores. Note that some layers contain parameters and other don't. In particular, the CONV/FC layers perform transformations that are a function of not only the activations in the input volume, but also of the parameters (the weights and biases of the neurons). On the other hand, the RELU/POOL layers will implement a fixed function. The parameters in the CONV/FC layers will be trained with gradient descent so that the class scores that the ConvNet computes are consistent with the labels in the training set for each image.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

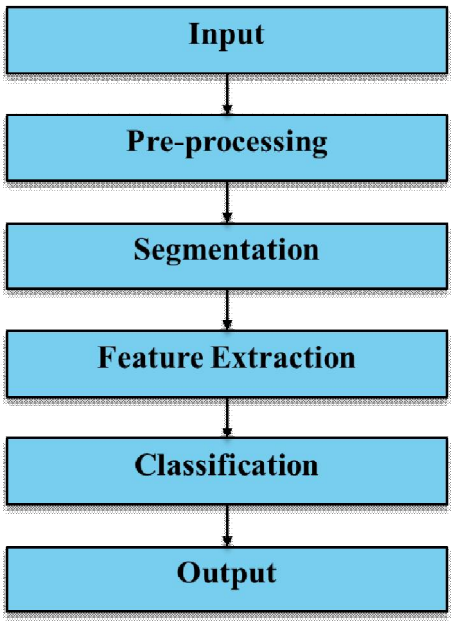
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

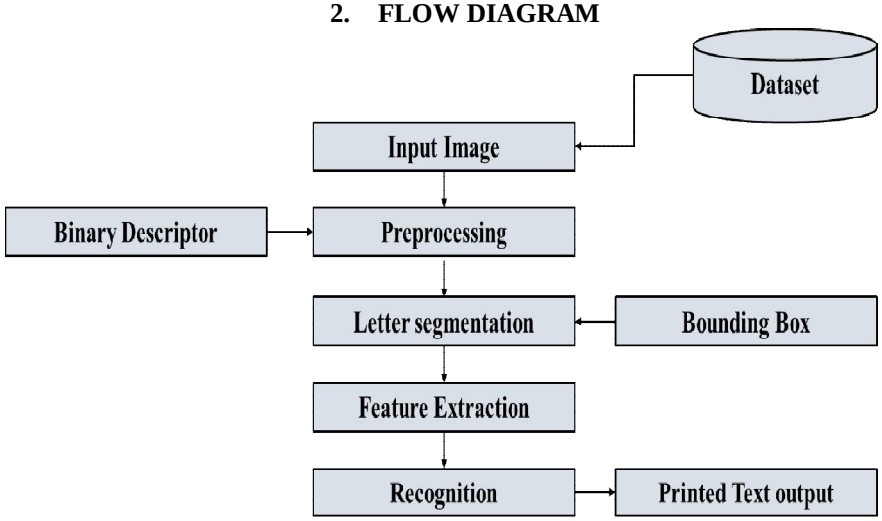
Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

1. BLOCK DIAGRAM



2. FLOW DIAGRAM



Explanation-

Preprocessing:

A binary image is a digital image that has only two possible values for each pixel. Typically, the two colors used for a binary image are black and white, though any two colors can be used. The color used for the object(s) in the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

image is the foreground color while the rest of the image is the background color. In the document-scanning industry, this is often referred to as "bi-tonal".

Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit—i.e., a 0 or 1. The names black-and-white, B&W, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel,such as grayscale images. In Photoshop parlance, a binary image is the same as an image in "Bitmap" mode.

Binary images often arise in digital image processing as masks or as the result of certain operations such as segmentation, thresholding, and dithering. Some input/output devices, such as laser printers, fax machines, and bi-level computer displays, can only handle bi-level images.

A binary image can be stored in memory as a bitmap, a packed array of bits. A 640×480 image requires 37.5 KiB of storage. Because of the small size of the image files, fax machine and document management solutions usually use this format. Most binary images also compress well with simple run-length compression schemes.

Binary images can be interpreted as subsets of the two-dimensional integer lattice Z2; the field of morphological image processing was largely inspired by this view.

Segmentation:

In geometry, the minimum or smallest bounding or enclosing box for a point set (S) in N dimensions is the box with the smallest measure (area, volume, or hyper volume in higher dimensions) within which all the points lie. When other kinds of measure are used, the minimum box is usually called accordingly, e.g., "minimum-perimeter bounding box". The minimum bounding box of a point set is the same as the minimum bounding box of its convex hull, a fact which may be used heuristically to speed up computation. The term "box"/"hyper rectangle" comes from its usage in the Cartesian coordinate system, where it is indeed visualized as a rectangle (two-dimensional case), rectangular parallelepiped (three-dimensional case), etc.

Feature Extraction:

Given an ambiguous image component, people can easily discriminate it in text or non-text, with much more informative knowledge about it, such as pixel-level text region segmentation and character information (e.g., ‘a’, ‘b’, ‘c’, etc.). Such low-level prior information is crucial for people to make a reliable decision. The text region segmentation allows people to accurately extract true text information from cluttered background, while the character label (assuming that people understand this language) helps them make a more confident decision on ambiguous cases.
Current text component filters are mostly learned with just binary text/non-text labels, which are insufficient to train a powerful classifier, making it neither robust nor discriminative. We wish to train a more powerful deep network for text task. We aim to ‘teach’ the model with more intuitive knowledge about the text or character. These important knowledge include lower-level properties of the text, such as explicit locations of text pixels and labels of characters.
Toward this, we propose the Text-CNN by training it with multi-level highly-supervised text information, including text region segmentation, characterlabel and binary text/non-text information. These additional supervised information would ‘tell’ our model with more specific features of the text, from low-level region segmentation to high-level binary classification. This allows our model to sequentially understand where, what and whether is the character, which is of great importance for making a reliable decision.
However, training a unified deep model that incorporates multi-level supervised information is non-trivial, because different level information have various learning difficulties and convergence rates. To tackle this problem, we formulate our training process as a multi-task learning (MTL) problem that treats the training of each task as an independent process by using one of the supervised information.

Recognition:

In this classification process In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

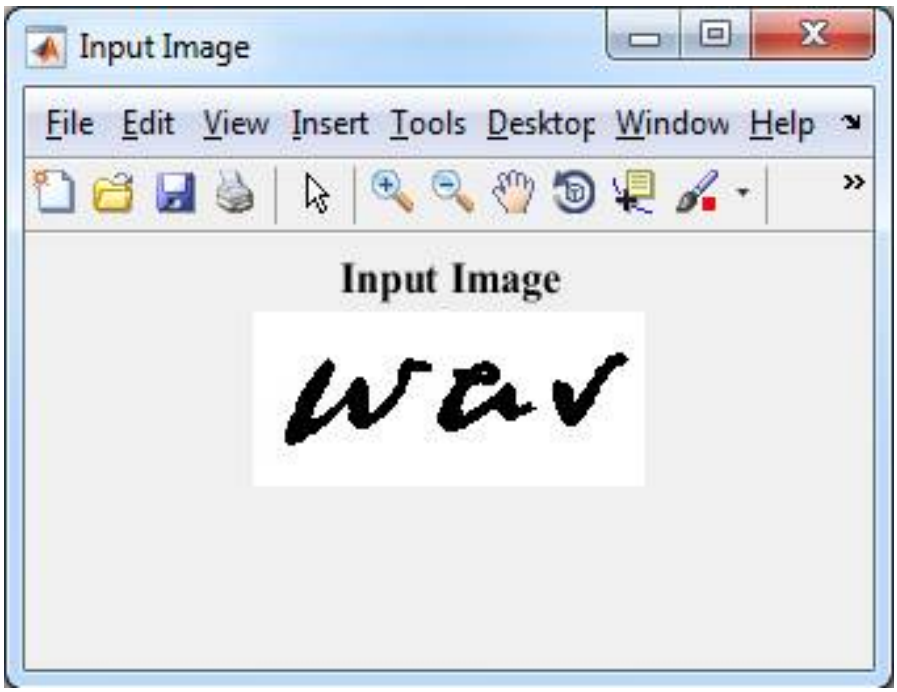
An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.
When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering and is often used in industrial applications either when data are not labeled or when only some data are labeled as a preprocessing for a classification pass.
Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to decide which class a new data point will be in. There are many hyperplanes that might classify the data. One reasonable choice as the best hyperplane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyperplane so that the distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum - margin hyperplane and the linear classifier it defines is known as a maximum margin classifier; or equivalently, the perceptron of optimal stability.

SVM classifier

(i) Data setup: our dataset contains three classes, each N samples. The data is 2D plot original data for visual inspection
(ii) SVM with linear kernel (-t 0). We want to find the best parameter value C using 2-fold cross validation (meaning use 1/2 data to train, the other 1/2 to test).
(iii) After finding the best parameter value for C, we train the entire data again using this parameter value
(iv) Plot support vectors
(v) Plot decision area

SVM maps input vectors to a higher dimensional vector space where an optimal hyper plane is constructed. Among the many hyper planes available, there is only one hyper plane that maximizes the distance between itself and the nearest data vectors of each category. This hyper plane which maximizes the margin is called the optimal separating hyper plane and the margin is defined as the sum of distances of the hyper plane to the closest training vectors of each category.

VI. RESULT



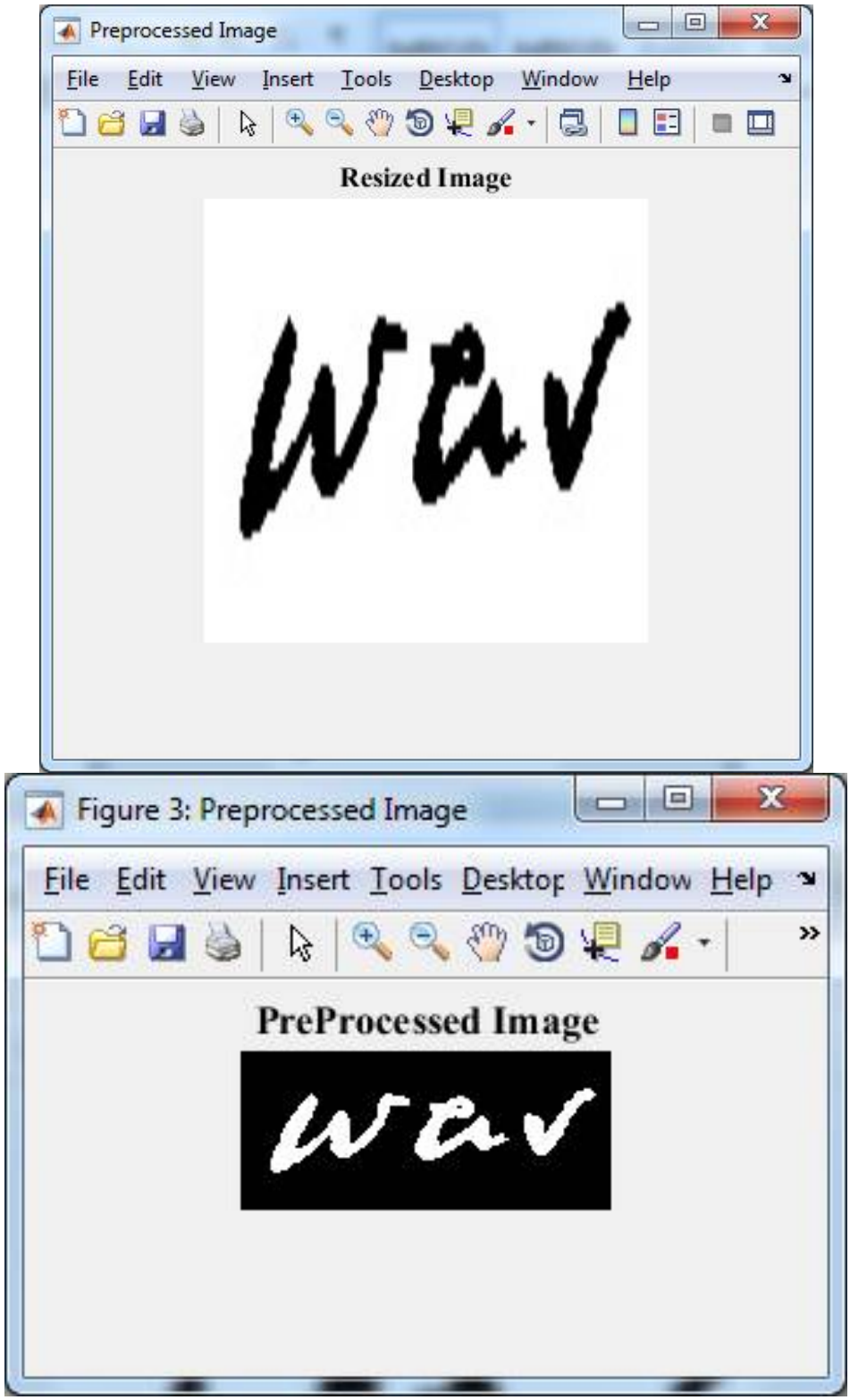
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018



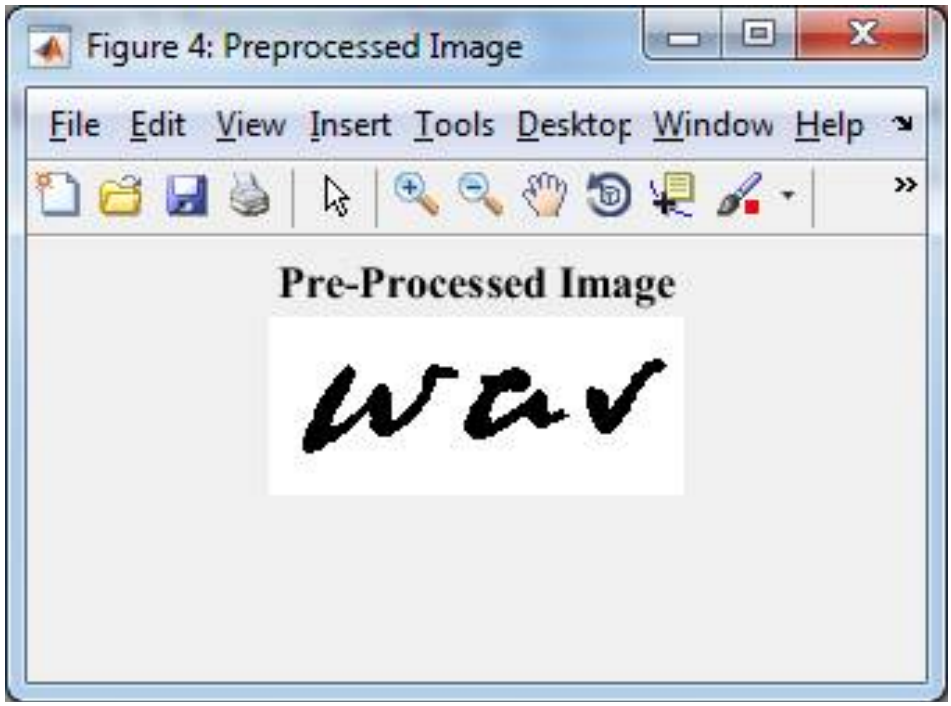
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

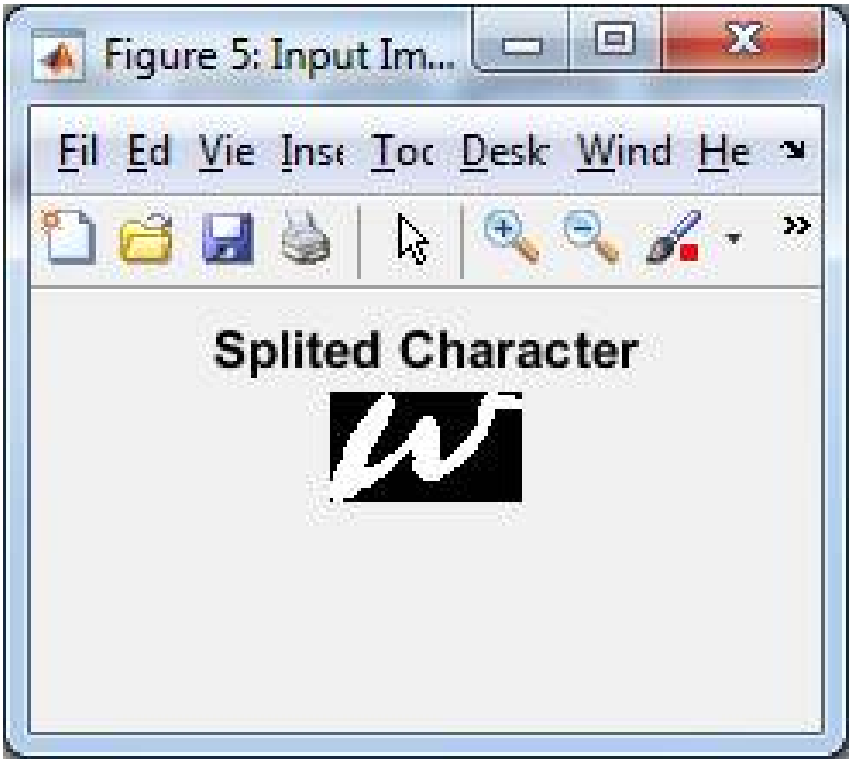
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018





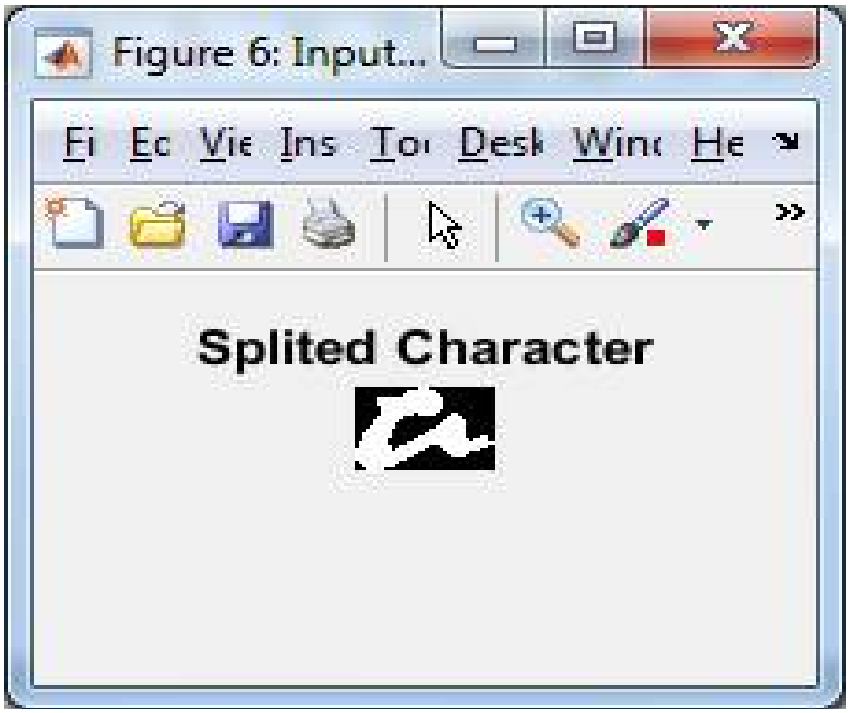
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

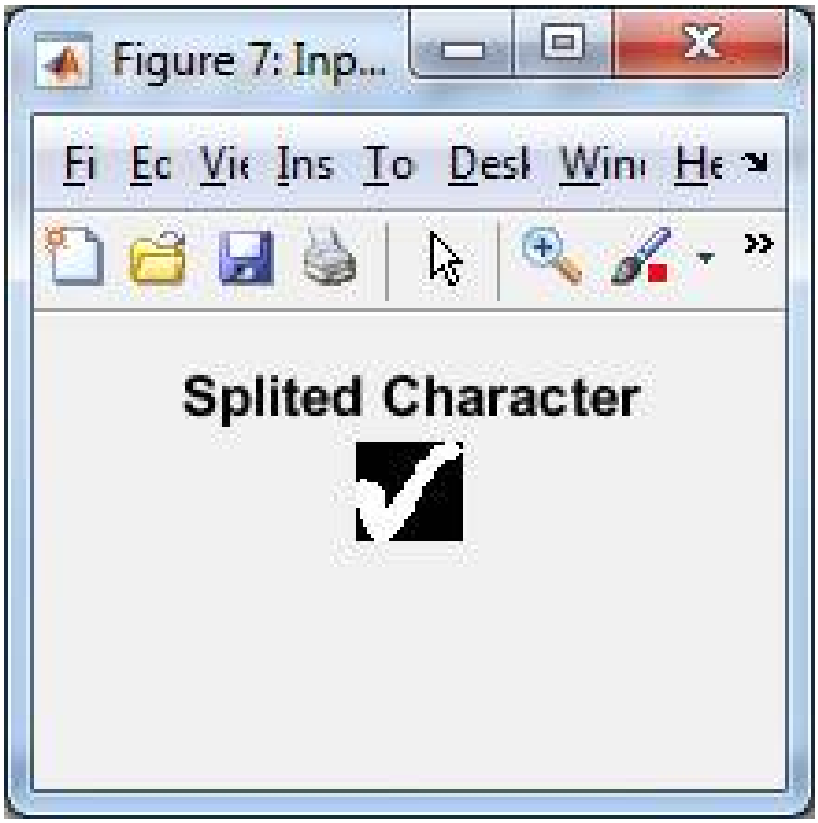
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018





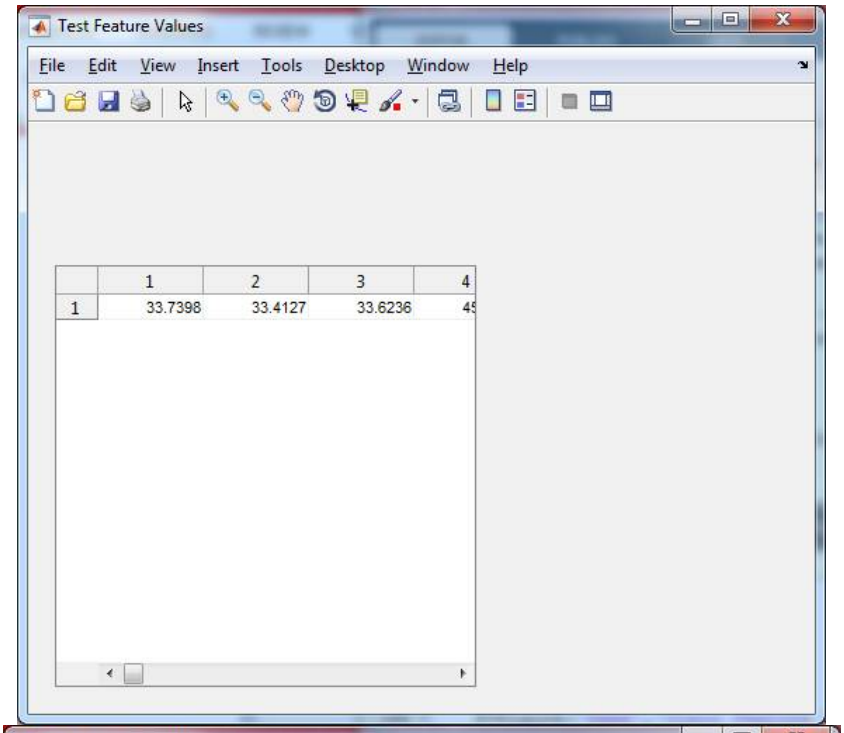
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

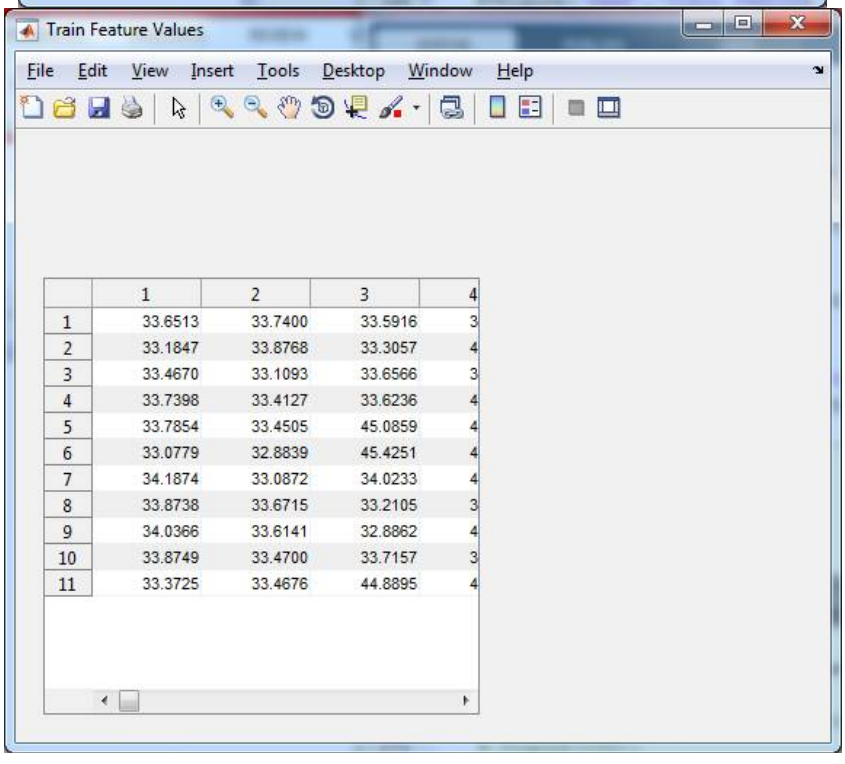
International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018





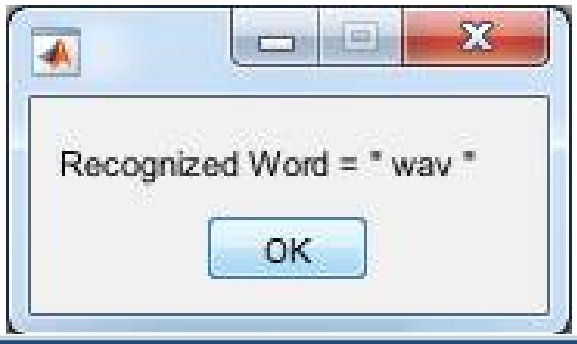
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018



Command Window

```
N =

     3


numbatches =

    12

epoch 1/1
Elapsed time is 0.585512 seconds.

numbatches =

    12

epoch 1/1
Elapsed time is 0.471978 seconds.

numbatches =

    12

epoch 1/1
Elapsed time is 0.470702 seconds.
Recognized Word = wav
```

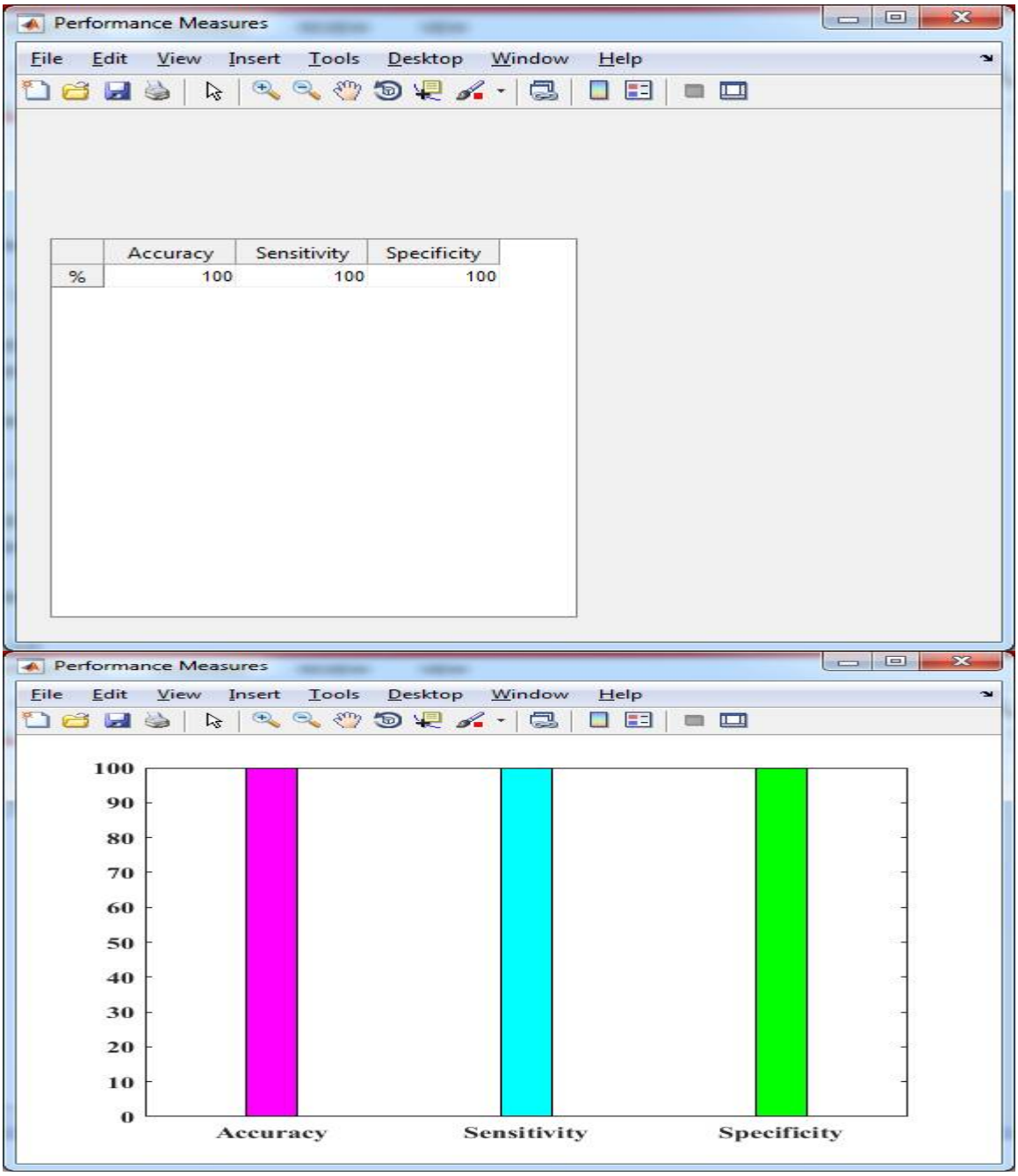
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018



VII. CONCLUSION

This process proposes an approach to represent and compare word images, both on document and on natural domains. We show how an attributes-based approach based on a pyramidal histogram of characters can be used to learn how to embed the word images and their textual transcriptions into a shared, more discriminative space, where the similarity between words is independent of the writing and font style, illumination, capture angle, etc. This attributes

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

representation leads to a unified representation of word images and strings, resulting in a method that allows one to perform either query-by-example or query-by-string searches, as well as image transcription, in a unified framework.

REFERENCES

[1] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Handwritten word spotting with corrected attributes," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 1017–1024.
[2] R. Manmatha and J. Rothfeder, "A scale space approach for automatically segmenting words from historical handwritten documents," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1212–1225, Aug. 2005.
[3] L. Neumann and J. Matas, "Real-time scene text localization and recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 3538–3545.
[4] L. Neumann and J. Matas, "Scene text localization and recognition with oriented stroke detection," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 97–104.
[5] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "PhotoOCR: Reading text in uncontrolled conditions," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 785–792.
[6] A. Fischer, A. Keller, V. Frinken, and H. Bunke, "HMM-based word spotting in handwritten documents using subword models," in Proc. 20th Int. Conf. Pattern Recog., 2010, pp. 3416–3419.
[7] V. Frinken, A. Fischer, R. Manmatha, and H. Bunke, "A novel word spotting method based on recurrent neural networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 2, pp. 211–224, Feb. 2012.
[8] R. Manmatha, C. Han, and E. M. Riseman, "Word spotting: A new approach to indexing handwriting," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., 1996, pp. 631–637.
[9] T. Rath, R. Manmatha, and V. Lavrenko, "A search engine for historical manuscript images," in Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2004, pp. 369–376.
[10] T. Rath and R. Manmatha, "Word spotting for historical documents," Int. J. Document Anal. Recog., vol. 9, pp. 139–152, 2007.
[11] J. A. Rodrıguez-Serrano and F. Perronnin, "Local gradient histogram features for word spotting in unconstrained handwritten documents," presented at the Int. Conf. Frontiers Handwriting Recognition, Montreal, QC, Canada, 2008.
[12] J. A. Rodrıguez-Serrano and F. Perronnin, "A model-based sequence similarity with application to handwritten wordspotting," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2108–2120, Nov. 2012.
[13] S. Espana-Bosquera, M. Castro-Bleda, J. Gorbe-Moya, and~ F. Zamora-Martinez, "Improving offline handwritten text recognition with hybrid HMM/ANN models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 4, pp. 767–779, Apr. 2011.
[14] I. Yalniz and R. Manmatha, "An efficient framework for searching text in noisy documents," in Proc. 10th IAPR Int. Workshop Document Anal. Syst., 2012, pp. 48–52.
[15] K. Wang, B. Babenko, and S. Belongie, "End-to-end scene text recognition," in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 1457–1464.
[16] A. Mishra, K. Alahari, and C. V. Jawahar, "Top-down and bottomup cues for scene text recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 2687–2694.
[17] J. A. Rodrıguez-Serrano and F. Perronnin, "Label embedding for text recognition," in Proc. Brit. Mach. Vis. Conf., 2013, pp. 5.1–5.12.
[18] C. Leslie, E. Eskin, and W. Noble, "The spectrum kernel: A string kernel for SVM protein classification," in Pacific Symp. Biocomput., 2002, pp. 564–575.
[19] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, "Text classification using string kernels," J. Mach. Learn. Res., vol. 2, pp. 419–444, 2002.
[20] H. Jegou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 1, pp. 117–128, Jan. 2011.
[21] S. Lu, L. Li, and C. L. Tan, "Document image retrieval through word shape coding," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 11, pp. 19313–1918, Nov. 2008.
[22] P. Keaton, H. Greenspan, and R. Goodman, "Keyword spotting for cursive document retrieval," in Proc. Workshop Document Image Anal., 1997, pp. 74–81.
[23] F. Perronnin and J. A. Rodrıguez-Serrano, "Fisher kernels for handwritten word-spotting," in Proc. 10th Int. Conf. Document Anal. Recog., 2009, pp. 106–110.
[24] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in Proc. Conf. Adv. Neural Inform. Process. Syst., 1999, pp. 487–493.
[25] B. Gatos and I. Pratikakis, "Segmentation-free word spotting in historical printed documents," in Proc. 10th Int. Conf. Data Anal. Recog., 2009, pp. 271–275.
[26] M. Rusinol, D. Aldavert, R. Toledo, and J. Llad~ os, "Browsing het- erogeneous document collections by a segmentation-free word spotting method," in Proc. Int. Conf. Data Anal. Recog., 2011, pp. 63–67.
[27] J. Almazan, A. Gordo, A. Fornes, and E. Valveny, "Efficient exemplar word spotting," in Proc. Brit. Mach. Vis. Conf., 2012, pp. 67.1–67.11.
[28] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog., 2005, pp. 886–893.
[29] T. Malisiewicz, A. Gupta, and A. Efros, "Ensemble of exemplarSVMs for object detection and beyond," in Proc. Int. Conf. Comput. Vis., 2011, pp. 89–96.
[30] A. Mishra, K. Alahari, and C. V. Jawahar, "Scene text recognition using higher order language priors," in Proc. Brit. Mach. Vis. Conf., 2012, pp. 127.1–127.11.
[31] A. Mishra, K. Alahari, and C. V. Jawahar, "Image retrieval using textual Cues," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 3040–3047.
[32] V. Goel, A. Mishra, K. Alahari, and C. V. Jawahar, "Whole is greater than sum of parts: Recognizing scene text words," in Proc. 12th Int. Conf. Document Anal. Recog., 2013, pp. 398–402.