MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011, Nara, JAPAN

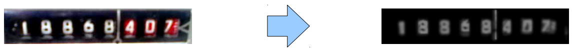
A Multi-Neural Network Approach to Image Detection and Segmentation of Gas Meter Counter

Angelo Nodari, Ignazio Gallo
University of Insubria, Dipartimento di Informatica e Comunicazione
via Mazzini 5, 21100 Varese, Italy
{angelo.nodari,ignazio.gallo}@uninsubria.it

Abstract

In this work we propose an automatic approach to detect and recognize text in natural images. Differently from other techniques which extract text from the whole image, our approach is trained to detect a particular area of interest where performs the search of text. Firstly we detect the region of interest and then we detect and segment the characters found. The proposed method may be used to read a generic text contained in a specific object from a single image. To show the results of our approach we focused on the problem of the automatic gas meter reading. The goal is to process a natural image containing a gas meter in order to extract only the information from the digits of the meter counter. Despite the complexity of the problem, our approach shows good results and the low computational time required by the algorithm permits to use it in real applications and in mobile devices.

(a) Region of interest detection



(b) Digits segmentation

Figure 1. Segmentation results produced by our algorithm. (a) An image of a meter and its segmentation obtained by our algorithm. (b) The image of the meter counter detected in the previous step and the respective digits segmentation obtained by a different configuration of our algorithm.

1 Introduction

Nowadays the detection and recognition of text from printed documents can be consider a solved problem. Consequently the research is focusing on the detection and recognition of text in natural images, as showed by the increasing number of related works [4, 2, 7, 1]. The main problems in this context are: poor image resolution due to a low-quality camera, blurred images and in particular motion blur caused by the shutter time of the camera, poor lighting or low contrast as a result of the overexposure, reflections, shadows etc... Due to the problems affecting the pictures in natural background, it is impossible to use standard document OCR techniques and for this reason it is necessary to develop alternative approaches. They can be divided in different categories according to [4]: Color-based Text Detection and Heuristic Connected Component Filtering; Edge-based Text Detection and Mathematical Morphology; Texture-based Text Detection. Our approach lies on a different category which uses a set of features as input of a neural network model and returns a soft map of the area of interest containing the digits of the meter counter.

There are several works concerning the problem of search a text in natural images and we report the most significant and closest to our approach. For example [1] uses a set of weak classifiers based on histogram, intensity and edge features selected by the Adaboost technique, [7] uses a graphical model to prune the false positive detected text in order to classify each pixel in a foreground or background class, [2] uses a bag of visual features Key Point detector and shape descriptors and requires high quality images to return good results. Differently from these works, that look for text in the whole image, our approach can be trained to detect and recognize a particular area of interest in which text should be found, ignoring the remaining part of the image. This is a useful feature in such applications that require to read written text above a specific object and in a defined position, like a license plate, a road sign etc. Moreover, due to the intrinsic characteristics of the proposed model, it is able to manage images with low resolution. Therefore, it may be used to recognize and read a generic text contained on a particular object in natural images using low power computing devices.

To show the result of our algorithm we use it to solve the automatic gas meter reading. These supply meters are used to measuring the consumption of gas, despite the existence of digital meters, that can acquire information in a digital way, their diffusion is slow and nowadays their reading is still done manually in many places. This operation is expensive and it requires a lot of time, for this reason we propose a real-time method to automatically read a gas meter using a single picture. Taking a picture is a very quick operation, moreover it can be stored and used as a guarantee of a correct reading and used in case of future controls. Moreover the low computational time required by the algorithm, makes it possible to use it on mobile devices. In this way it is possible to provide a mobile device to the people in charge of reading the meter. In addition to the text detection issues there are other problems afflicting this application such as: different digit fonts, different type of supply meters concerning size, shape, color etc... To solve this problem, our approach is divided in two stages: firstly we detect the region of interest in where the digits appear and than we segment the digits found. For completeness, a test using an open-source Optical Character Recognition (OCR) algorithm was done in order to recognize the digits singularly. The ultimate aim is to measure the ability of the available tools in real and difficult scenarios, as that taken into consideration.

2 Proposed Method

We propose a new segmentation system based on a set of neural networks, for the automatic number recognition of a gas meter, from now on called MultiNOD.

We chose to use a model based on supervised neural networks for two main reason: to overcome many of the problems that afflicts the images in a real scenario and because neural models have a high generalization ability and good robustness.

The MultiNOD model consists of a tree of neural networks, where each network uses a sliding window of size W_S to read the features extracted from the input image and then resized according to a second parameter I_S, which identify the smaller side of the resized input image. So each node $C^n_{I_S,W_S}(F_1,\ldots,F_m)$ or neural network, has two parameters (I_S, W_S) and a set of features $(F_1,\ldots,F_m)$ from which to construct the patterns. A node produces in output a soft classification map (segmentation map) where the pixels containing higher values identify the objects of interest (see the two examples showed in Figure 1). In the present work, the structure of a single node consists in a feed-forward Multi-Layer Perceptron (MLP), trained using the learning algorithm Rprop (Resilient Backpropagation) proposed by Riedmiller and Braun [6]. But we want to stress that, MultiNOD can use other learning paradigms as an alternative to MLP, even different types of nodes for a given configuration. For example, the MLP model can be replaced by a SVM, RBF, Decision Tree, etc., or a combination of these.

The particular aspect of this model lies in the connection between nodes, which in practice means that the output of a node becomes the input of a parent node. This means that some features of $(F_1,\ldots,F_m)$ could be the output of child nodes. A node, with one or more segmentation maps in input, uses these for two different reasons: as confirmation of the true positives detections and as an instrument for a multi-scale approach. In practice, from any segmentation map F_i all the bounding rectangles (ROI) of the contours retrieved from the binarized F_i map are treated as a new image. In this multiscale scenario, the output segmentation map is obtained as the average of all the maps obtained from each ROI.

The set of features, such as information on the edges, color, etc., can be directly extracted from the input images. In this work we used the blue (B) and green (G) channels, and the normalized red channel ($RN = R/(R+G+B)$), as color features. Information about edges were extracted using a 1-D vertical (EV_s) and horizontal (EH_s) derivative filter, using the mask $[-1, 0, 1]$. Another filter that was used to extract information from the edges is the Sobel filter (Sob_s). The s parameter identifies the scale, for example 0.5 means that the size of the image has been halved before calculating the feature. More information on edges were extracted using some Harr-like features proposed by Viola and Jones in [9]. In particular the difference between the sum of the pixels within two rectangular regions, defined as *two-rectangle feature*, was used. The regions have the same size and shape and are horizontally ($H2H_{w\times h}$) or vertically ($H2V_{w\times h}$) adjacent. The two parameters w and h identify the width and height of the rectangles, respectively.

Figure 2. Some examples of images from the dataset used in this work. Many of these images are troublesome due to occlusions, lighting conditions, blur, perspective distortion, etc..

The MultiNOD configuration was chosen considering which nodes in which layers give better results using a trial and error procedure. We have used the following strategy to optimize the results: nodes belonging to the first level of the tree are used to find out areas of interest, and nodes of subsequent layers are used to eliminate the false positives and to confirm the true positives, taking advantage of the multiscale strategy described above.

The MultiNOD system was applied twice, once to detect the meter region and a second time to identify the single digits.

3 Experiments

Two experiments were performed using the proposed model: a first experiment to search an acceptable configuration to solve the problem of finding the digits of the meter counter, and a second experiment to understand how to configure the MultiNOD model in order to segment the digits. The second experiment is essential to the OCR phase because many algorithms require accurate detection of the area containing the characters in order to well recognize the character.

Tests were performed using a single thread C# code, on an Intel®Core™Duo processor T8100 @2.10Ghz. The two phases have created an output map in an average time of 500 ms per image.

To assess the detection system, precision (P) and recall (R) have been used:

$$P = \frac{correct}{actual}, \quad R = \frac{correct}{possible} \tag{1}$$

stating that *correct* is the number of objects correctly detected by the system, *actual* is the total number of objects recognized by the system, *possible* is the total number of objects we expected from the system. A measure that usually is used to summarize Precision and Recall values is the F-measure $F = 1/(\lambda P + (1 -$

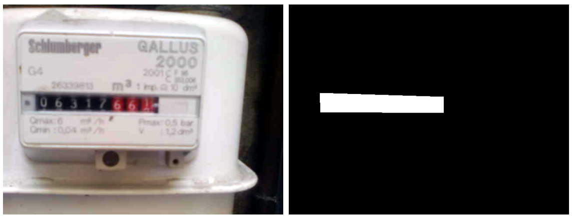


Figure 3. Input/output training example for the meter detection problem. Only the pixels belonging to the meter has been labeled with as foreground, while all others were labeled as background.

$\lambda)R)$ where usually $\lambda = 0.5$. Average Precision (AP) is another measure that approximates the area under the PR-curve and that, for its readability, is usually used to compare different object detection algorithms.

To assess the segmentation system, the intersection/union of class labels (*Acc.*), computed over all pixels, was used: $Acc. = tp/(fn+fp+tp)$. This metric is interesting because penalizes both over- and underestimates objects. The overall accuracy is the average between the class and background accuracy.

For more details on these measures and how an object is considered correctly detected or segmented see [3].

3.1 Counter detection

The dataset *Gas-Meter-Counter*, used in the first experiment, consists of 102 training images and 51 testing images. All images were acquired by different operators using different typology of devices. Figure 2 shows some typical examples of this dataset and it is evident that, under these conditions, the detection and segmentation problem becomes very difficult. From the output image shown on the right side of Figure 3, it is clear that the objective of the MultiNOD model is to learn to identify the area containing the counter. This is not a trivial problem because a lot of written visible in the image can be confused with the numbers of the meter.

The MultiNOD configuration used here is shown in Figure 4. Nine different MLP neural networks $C^n_{I_S,W_S}(F_1,\ldots,F_m)$ were used, configured with some of the set of predefined features. The parameters I_S and W_S and the set of features $(F_1,\ldots,F_m)$, have been specifically selected for each node to highlight some typical features of the meter area to be detected. For example small values of I_S and values of W_S, very close to the image size I_S (such as for nodes C^8 and C^9 in Figure 4), create an output map with a relatively high precision but, being a very low-resolution map, provides an idea of the location of the area of interest with a very low recall. Conversely, great I_S values and relatively very small W_S increase the recall measure but reduce the precision (see for example nodes C^6 and C^7 in Figure 4).

The results in terms of detection and segmentation are shown in Table 1. The high recall is usually a desirable goal if we want to recognize the sequence of digits from each meter. In terms of segmentation, it is important to note that more than 95% of pixels were correctly classified as background, and this is certainly

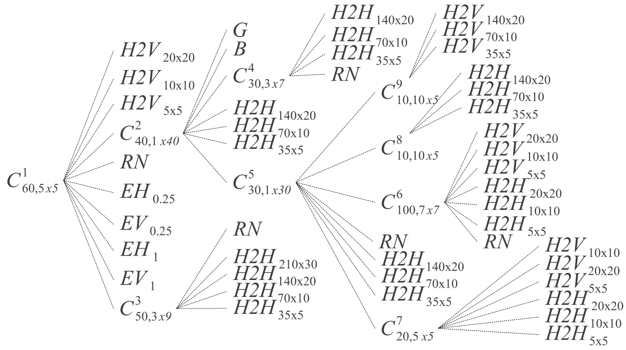


Figure 4. MultiNOD configuration used for the meter detection. Each node $C^n_{I_S,W_S}$ is representative of a neural network trained on the training set built from all the features that receive input. All other nodes are simple features extracted from input images.

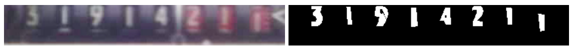


Figure 5. Input/output training example for the digit segmentation problem. The figure-ground segmentation map has been segmented by hand to train the model to recognize and segment only the digits of the meter.

an important result, to avoid detection of many false positives.

Table 1. Detection and segmentation results obtained on the test set *Gas-Meter-Counter* for the meter detection problem.

Detection		Segmentation	
Precision	63%	Acc. background	95.5%
Recall	78%	Acc. object	51.7%
F-Measure	70%	Average Acc.	73.6%
AP	50.4%		

3.2 Digit segmentation

For the digit segmentation problem a second dataset was constructed, consisting of 47 training images and 47 testing images. All images were cropped from original *Gas-Meter-Counter* dataset. As shown in the training example of Figure 5, in this experiment the MultiNOD model has to learn to detect and segment the digits of the meter counter presented as an input image.

The MultiNOD configuration, used in this second experiment, is composed of eight different neural networks and is shown in Figure 6. Nodes C^7 and C^8 identify the image areas containing the digits, while all subsequent nodes try to improve the segmentation for each digit.

The detection and segmentation results are shown in Table 2. Precision and recall are acceptable and have the same order of magnitude, while the digits segmentation accuracy of the test set is close to 50%. The very high accuracy of background assures us that many problems due to noise present in the proximity of

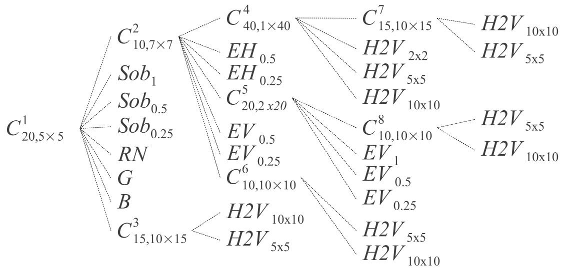


Figure 6. MultiNOD configuration used for digit segmentation. Each node $C^n_{I_S,W_S}$ is representative of a neural network trained on the training set built from all the features that receive input. All other nodes are simple features extracted from input images.

the digits have been removed. As we can see from the segmentation example shown in Figure 1(a), results of this second step help the final OCR phase to reduce the number of false positives.

Table 2. Digit detection and segmentation results.

Detection		Segmentation	
Precision	68%	Acc. background	92.8%
Recall	79%	Acc. object	49.5%
F-Measure	73%	Average Acc.	71.6%
AP	55.8%		

3.3 Digit reading

A test using an open-source OCR algorithm was done in order to measure the ability of a "standard tool" in the real and difficult scenarios taken into consideration.

The output of the segmentation phase identifies the area of interest containing a single digit and then an OCR algorithm is used to read this character. The OCR algorithm is configured to recognize only 10 classes of characters, in particular only the digits from 0 to 9. We used Tesseract [8], an open source OCR engine, to read the digit obtaining the following result on a dataset of 332 digits correctly detected by the segmentation phase: 58,73%. This result highlight the complexity of the OCR task in this domain, due to some recurring problems listed above (see some examples in Figure 7). Advanced recognition techniques can be used in place of the standard OCR algorithm used in this work to improve the results, such as some interesting intelligent character recognition systems based on artificial neural networks like that explained in [5].

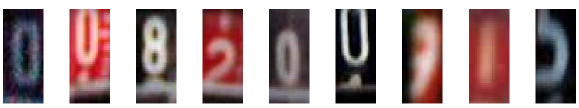


Figure 7. Some examples of difficult digits to be read. Very noisy images, low contrast, distinguishing marks too close to the digit, occluded or too blurry numbers.

4 Conclusion

In this work we have presented an approach to detect and subsequently segment digits of a gas meter counter in natural images. The proposed model is based on a set of neural networks that work on low-resolution images, and give a result in about half a second per image. The results are promising and, for example, ensure the feasibility of an application for mobile devices able to automatically read the numerical value of a meter. The main feature of the MultiNOD model, shown in this study, is the detection and segmentation of text in a specific area of interest. This task is performed in a single step without the needed of any preprocessing or postprocessing phases. Moreover this strategy can be applied on different applications with a low conversion effort, consequently we are investigating how to use this approach in different domains.

References

[1] Xiangrong Chen and A.L. Yuille. Detecting and reading text in natural scenes. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–366 – II–373 Vol.2, 2004.

[2] T. E. de Campos, B. R. Babu, and M. Varma. Character recognition in natural images. In *Proceedings of the International Conference on Computer Vision Theory and Applications, Lisbon, Portugal*, February 2009.

[3] Mark Everingham, Andrew Zisserman, Christopher K. I. Williams, Luc Van Gool, Moray Allan, Christopher M. Bishop, Olivier Chapelle, Navneet Dalal, Thomas Deselaers, Gyuri Dork, Stefan Duffner, Jan Eichhorn, Jason D. R. Farquhar, Mario Fritz, Christophe Garcia, Tom Griffiths, Frederic Jurie, Daniel Keysers, Markus Koskela, Jorma Laaksonen, Diane Larlus, Bastian Leibe, Hongying Meng, Hermann Ney, Bernt Schiele, Cordelia Schmid, Edgar Seemann, John Shawe-taylor, Amos Storkey, Or Szedmak, Bill Triggs, Ilkay Ulusoy, Ville Viitaniemi, and Jianguo Zhang. The 2005 pascal visual object classes challenge. In *Selected Proceedings of the First PASCAL Challenges Workshop*, 2005.

[4] Keechul Jung, Kwang In Kim, and Anil K. Jain. Text information extraction in images and video: a survey. *Pattern Recognition*, 37(5):977 – 997, 2004.

[5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. volume 86, pages 2278–2324, November 1998.

[6] Martin Riedmiller and Heinrich Braun. A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *IEEE International Conference on Neural Networks*, pages 586–591, 1993.

[7] Huiying Shen and J. Coughlan. Finding text in natural scenes by figure-ground segmentation. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 113 –118, 2006.

[8] R. Smith. An overview of the tesseract ocr engine. In *Proceedings of the Ninth International Conference on Document Analysis and Recognition - Volume 02*, pages 629–633, Washington, DC, USA, 2007. IEEE Computer Society.

[9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 2001.