

TEXT RECOGNITION FOR OBJECTS IDENTIFICATION IN THE INDUSTRY

Pedro M. B. Torres

Instituto Politécnico de Castelo Branco, Castelo Branco, Portugal

pedrotorres@ipcb.pt

Abstract. In line with 4th industrial revolution (Industry 4.0), the mechatronics and related areas are fundamental to boost the developments of industry digitalization. However, it should not be forgotten that artificial intelligence (AI) has a great preponderance on the development of autonomous and intelligent systems incorporating the advances in mechatronics systems. It is common in different industries the need to identify and recognize products or objects for different purposes such as counts, quality control, selection of objects, among others. For these reasons, pattern recognition is increasingly being used in systems on the shop floor, usually implemented in computer vision systems with image processing in real time. This work focuses on automatic detection and text recognition in unstructured images for use on shop floor mechatronic systems with vision systems, to identify and recognize patterns in products. Unstructured images are images that does not have a pre-defined image model or is not organized in a predefined manner. Which means that there is no predefined calibration model, the system must identify and learn by itself to recognize the text patterns. The techniques of character recognition, also known as OCR (Optical Character Reader), are not new in the industry, however the use of machine learning algorithms together with the existing techniques of OCR, allow endow the systems of greater intelligence in the patterns recognition. The results achieved throughout the paper, demonstrates that it is possible to identify and recognize text in objects based on unstructured images with a high level of accuracy and that these algorithms can be used in real time applications.

Keywords: Industry 4.0, Mechatronics, Machine Vision, Machine Learning, Text Recognition.

1. Introduction

Industry 4.0 is the current trend for industry digitalization. It involves technologies such as advanced robotics, artificial intelligence, sophisticated sensors, cloud computing, internet of things (iot), data analytics, digital fabrication, cyber-physical systems, communications, mobile devices, etc. These technologies are joined together to integrate the physical and virtual worlds. Under the paradigm of industry 4.0 it is intended constructed the factories of the future, so called smart factories. Over the industrial internet of things (iiot), iot for industry, cyber-physical systems monitor and control the physical processes, communicate and cooperate with each other and with humans in real time. The Industry 4.0 movement is first proposed during the Hanover Fair 2011 in Germany and in April 2013 is writing the "Recommendation for implementing the strategic initiative INDUSTRIE 4.0" prepared for the German Federal Ministry of Education and Research that reports the vision for Industry 4.0 [1].

Automation and robotics needs accompanying these new production trends and off course also

includes machine vision systems for robust identification processes and a high detection rate [2]. In industry is very usually the existence of machine vision systems to provide automatic image-based inspection and analysis for applications such as automated inspection, process control and robot orientation. The combination with machine learning techniques helps to improve the machine vision systems, to identify and recognize patterns based on computer vision techniques. Computer vision tasks include methods for acquiring, processing, analysing and understanding scenes in images. The extraction of high-dimensional data from the real world is important to produce information that assist the mechatronics systems in the decision processes. The object recognition and the text recognition are a sub-domains of computer vision techniques. For other hand, machine learning is field that gives "computers the ability to learn without being explicitly programmed". It involves the study of pattern recognition, computational learning, artificial intelligence, to study and construct algorithms that can learn from and make predictions on data.

Optical character recognition (OCR) is an application example of machine learning techniques. OCR is a technic used by machine vision systems to

convert scanned images of text to electronic text so that digitized data can be searched, indexed and retrieved. OCR engines are developed and optimized not only for the industry but for multiple real world applications. Each application requires processing a lot of image datasets to train and optimize the algorithms. The accuracy of the results depends on the efficiency of the training process and is based on continuing learning that the systems can be smarter. There are many different approaches to solving the OCR problem. Neural networks (NN) are one of the most common and popular approaches used.

Artificial Neural Networks (ANNs) can be understood as a computational model used in machine learning that incorporates classifiers that can help solve OCR-type problems [3]. The ANN is an information processing paradigm inspired in human brain processes. Are sets of mathematical models that represent some of the observed properties from biological nerve systems and draw on the analogies of adaptive biological learning. The neural network topology consists of many highly-interconnected processing elements (neurons) and involves synaptic connections that exist between neurons.

A popular and simple NN approach to the OCR problem is based on feed forward NN with backpropagation learning. This approach consists on that, first we need to prepare a training set and then train an NN to recognize patterns from the training set. In the training step, we teach the network to respond with the desired output for a specified input. Each training sample is represented by two components, possible input and the desired network's output given that input. After concluded the training step, we can give an arbitrary input to the network that will form an output, from which we can resolve a pattern type presented to the network. In [4] it is explained with an example how neural networks works for OCR learning.

The problem with text recognition in images is that it is highly sensitive to any change. There are important factors during the acquisition process such as illumination, skewing and image rotation that can affect the quality of features extraction. The Optical Character Recognition typically involves the extraction of character regions from an image, segmentation of the image into text and background, conditioning of the images for noise remove for example, segmentation of characters, normalisation of character size, feature detection, classification and verification.

In the literature, there are a series of research works that addresses the recognition of text in images. In [5] the authors propose a novel text detection algorithm, which employs edge-enhanced Maximally Stable Extremal Regions (MSER) as basic letter candidates.

To overcome the sensitivity of MSER with respect to image blur and to detect even very small letters, an edge-enhanced MSER which exploits the complimentary properties of MSER and Canny edges is developed. In [6] the authors present a new method to locate text in images with complex background. It combines efficiently MSER and a locally adaptive thresholding method. The result is a connected-component-based approach that extracts basic letter candidates using a series of easy and fast-to-compute features. In [7] a novel Connected Component based methodology for text detection in natural scene images is presented. MSERs are first utilized as potential text regions. In [8] the authors propose an end-to-end real-time text localization and recognition method.

In this work, will be presented the results obtained in the process of detection regions in natural images that contain text, very important in the shop floor to recognize objects in the production and distribution lines. It is used the MSER feature detector, described in section 2, for finding text regions.

Figure 1 shows the machine vision system configuration overview idealized for our implementation. Typically, the system can be composed for sensors to detect object, robot arms for manipulation, PLCs for controlling the automated cell, and a vision system for identifying the objects transported in different conveyors. It is common the use of a SCADA system for monitor and supervisor, and in line with there are the perspectives for the cyber-physical systems, the use of remote devices to interact with the automated and robotized cells. The text recognition algorithms run on the workstation in the shop floor or in the cloud.

The remaining of the paper is organized as follows: Section II describes the methods and procedures; Section III presents the experimental results. Finally, the conclusions and future work are given in section IV.

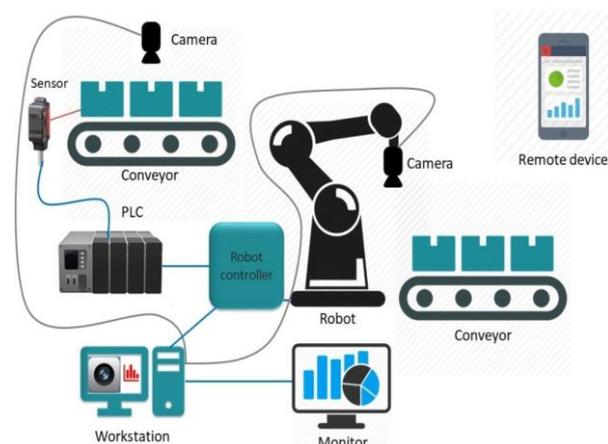


Fig. 1. Setup overview of the machine vision system

2. Methods and Procedures

This section describes the followed methodology to identify and finding text in image regions considering images not pre-processed, commonly named unstructured images.

To understand the steps involved in recognizing text, Figure 2 shows the general OCR pipeline with the different tasks. From an unstructured image, you need to detect potential regions with text, but before text detection, in most cases you need to pre-process images with filters and convert grayscale images to

improve results. Sometimes different regions are identified as a potential area with text but have no text, so it is important to identify and eliminate those regions. The next task is character targeting to identify each character in each detected text region.

After this task, the characters are recognized by a training process where they usually involve machine learning techniques like ANNs to recognize each character presented in each region. The final task is to unite all the recognized characters to form the final words.

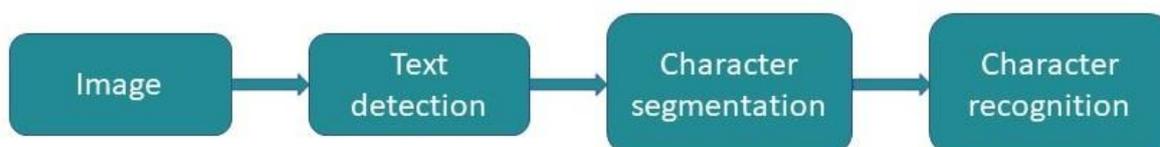


Fig. 2. OCR pipeline

MSER is a well-known method for detecting regions in images. The algorithm extracts from an image several co-variant regions, called MSERs. An MSER is a stable connected component of some gray-level sets of the image. The MSER is based on the idea of taking regions that remain almost equal across a wide range of thresholds. All pixels below a given threshold are white and all pixels above or equal are black. If a sequence of images with thresholds is shown, with the frame t corresponding to the threshold t , we will first see a black image,

then the white spots corresponding to the minimums of local intensity will appear. These white dots will eventually fuse until the entire image turns white. The set of all components connected in the sequence is the set of all the end regions. Optionally, elliptical frames are attached to the MSERs by fitting ellipses to the regions. Those regions descriptors are kept as features.

Table 1 presents the steps followed to recognize the text present in the images, based on MSER detector.

Table 1. Methodology followed in this work:

- Phase 1
 - Load images;
 - Convert images to gray-scale;
 - Detect MSER regions;
- Phase 2
 - Measure geometric properties of image regions;
 - Remove non-text regions based on geometric properties;
- Phase 3
 - Merge text regions into words or text lines;
- Phase 4
 - Recognize text within each bounding box based on OCR methods;

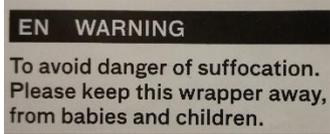
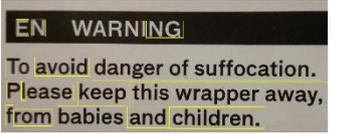
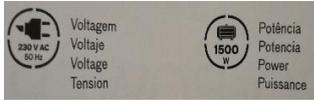
3. Experimental Results

Experimental tests were performed on different types of real unstructured images, common in industry. Bar code images and images with important technical information to be detected by the automated systems were considered.

In table 2 is showed results obtained from those images. As can be seen the results depend on the images quality, however in the most cases, it is recognized most of the text presented in the images.

Even in images as logos with small letters it is possible to automatically detect and recognize the present text.

Table 2. Experimental results obtained from different types of images

Original Image	Regions with text detected	Output
		B3fee 1 6a9bfe
		To ease rom babies and children Avoid Keep this wrapper away, danger of suffocation. IARN NG
		AC Votagem Voltage Tension Puissance Power Potencia Potência 1500
		in Malaysia Made PP3 PE

4. Conclusions

In this work, MSER algorithm is used for text detection in unstructured images, commons in the industry. MSER is very sensitive with blur images and small letters, however it reveals to have potential so that associated with other image processing techniques such as edge detectors can improve the results. The results are promising, considering that the input images can be from different fonts without a predefined pre-processing.

Working on this framework and considering the new trends of machine learning and deep learning applied to text recognition, the results will be significantly improved in future. As future work, it is expected to improve the performance of results by considering new methods to identify and remove non-text regions and test different machine learning methods to recognize the words.

References

- [1] Henning, K., Wolfgang, W., Johannes, H.: Recommendations for Implementing the Strategic Initiative INDUSTRIE 4.0: Securing the Future of German Manufacturing Industry; Final Report of the Industrie 4.0 Working Group, Forschungsunion (2013).
- [2] <http://www.qualitymag.com/articles/93296-machine-vision-in-iiot>. accessed at: 26.04.2017.
- [3] Mehta, H., Singla, S. and Mahajan, A.: "Optical character recognition (OCR) system for Roman script & English language using Artificial Neural Network (ANN) classifier," *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*, Bangalore, pp. 1-5 (2016).
- [4] <http://www.cvisiontech.com/resources/ocr-primer/ocr-neural-networks-and-other-machine-learning-techniques.html>. accessed at: 04.05.2017.
- [5] Chen, H., Tsai, S.S., Schroth, G., Chen, D.M., Grzeszczuk, R. and Girod, B.: "Robust Text Detection In Natural Images With Edge-Enhanced Maximally Stable Extremal Regions," in 18th IEEE International Conference on Image Processing (2011).
- [6] González, Á., J. J. Y. Luis Bergasa, M. e Bronte, S.: "Text Location in Complex Images," in 21st International Conference on Pattern Recognition, Tsukuba, Japan (2012).
- [7] Li, Y. and Lu, H.: "Scene Text Detection via Stroke Width," in 21st International Conference on Pattern Recognition, Tsukuba, Japan (2012).
- [8] Neumann, L. and Matas, J.: "Real-Time Scene Text Localization and Recognition," in 25th IEEE Conference on Computer Vision and Pattern Recognition, USA (2012).