Document image analysis by a mobile robot for autonomous indoor navigation

Dalia Marcela Rojas Castro, Arnaud Revel, Michel Ménard

Document Image Analysis by a Mobile Robot for Autonomous Indoor Navigation

D.M. Rojas Castro, A. Revel, M.Ménard
Computer Science Department, Laboratory L3i of La Rochelle University
Avenue Michel Crépeau 17042, La Rochelle, France
{dalia_marcela.rojas_castro, arnaud.revel, michel.ménard} @ univ-lr.fr

Abstract—**This paper proposes the analysis of a paper-based floor plan of a building for reference purposes in the navigation task of a mobile robot. The novelty of this approach lies in the combination of this information extracted from the document image in real time just prior to real world exploration with dynamic visual perception. An artificial neural system integrates both sets of information by having the robot recognize navigation signs from the floor plan and then look for them in the expected sequence in the environment. Thereafter, while determining and executing appropriate movements as a response to the environment, the robot links and learns to associate the undertaken movements to the respective signs. The analysis of the floor plan involves a detailed process permitting the robot to extract the required information in real time. This process is elaborated in this paper, along with experiments with the robot using the proposed architecture in a simple navigation scenario, hence proving the feasibility of the approach.**

Keywords—floor plan analysis; symbol recognition; robot navigation; document information merging.

I. INTRODUCTION

Digital floor plans of buildings (such as hospitals, schools, residential complexes or factories) are typically not as readily available as digital maps of entire cities, countries etc. Hence, when trying to navigate an unknown building, one has to rely on floor plans available in physical form at the entrance of the building or on paper to achieve one's final destination. Since such a floor plan provides one of the fastest way to access comprehensive information about the inside of the building, it can be used as *a priori* information for reference purposes in the navigation task of a mobile robot.

This paper proposes a document analysis-based system allowing a robot to emulate the cognition process of a human brain when navigating an unknown building by "reading" a map or a floor plan using its camera and "remembering" a sequence of navigation signs. These memorized signs will serve as reference landmarks in the real environment allowing it to navigate towards its final destination.

Navigation strategies allowing mobile robots to autonomously travel from a starting point to a goal are extremely diverse [1]. Control architectures are the essential component of such navigation. They define the capacities of the robot to plan a path trajectory, to undertake autonomous decision-making and to execute the appropriate reaction according to the perceived environment information.

Therefore, the underlying idea of this work is to present the conception of such system, based on an approach at the same time top down (use of an a priori model of the environment represented in a paper-based document) and bottom up (reactive navigation based on the perceived visual information of the real environment), that allows the robot to achieve its goal while overcoming the challenges presented by an environment prone to change. Hence, a document analysis-based control architecture which integrates both *à priori* and dynamic visual perception has been conceived. This integration is brought about using a neural network system.

Our approach addresses the control in a manner that is significantly different from existing robot navigation systems in two ways. Firstly, the *a priori* global knowledge of the environment is gathered by the robot in real time from a paper-based floor plan of the building and it is used to only corroborate the dynamic visual information perceived during navigation instead of directly controlling the actions of the robot as found in the literature [2]. Secondly, instead of using a complete motion path, this architecture makes use of navigation signs (and their expected sequence in the route) extracted from the said document, the associated directional implications of which may be learnt as a consequence of a stimulus-response model during navigation. As a result of this perceptual combination in real time, the robot is capable of performing autonomous navigation, memorizing and learning sensory-motor associations, decision-making and rapidly responding in dynamic environments.

The reminder of this paper is organized as follows. Section II introduces the document analysis based control architecture and describes the integration of the two types of perceptual information mentioned above using a neural network. This approach is the main contribution of this paper. Section III describes in detail the specific image processing steps allowing the extraction of the navigation signs from a document in real time. Results of experiments implementing the system in a humanoid robot (NAO) are described in section IV, and section V presents conclusions and potential future directions.

II. DOCUMENT ANALYSIS BASED NAVIGATION SYSTEM

A. Context

The global knowledge of the world is represented by a paper-based floor plan that is placed in front of the robot's camera just once, before the navigation activity starts (Fig. 1a). It contains the important information to define a potential

navigation trajectory that might eventually be modified depending on the unforeseen situations presented in the environment (e.g. obstruction of the path by a collapsed wall). In this work, this information is represented by navigation signs used as reference points that are expected to be seen by the robot in the real world navigation (black circles with specific white patterns drawn within, associated with left and right directions as shown in Fig 1b). By means of computer vision methods, the robot "reads" the floor plan, generates an optimal plan to reach the goal, recognizes the navigation signs within the path, extracts and "memorizes" the sequence of signs arranged from the closest point to the furthest with respect to the starting point (see section (III)).

The robot then integrates this extracted sequence of signs into the neural system Fig. 2 and combines it with the visual perception of the environment in real-time (Fig 1c). The neural system uses a cognitive mechanism performing recognition of a particular sign and allows learning of its association to the realization of a particular action during navigation. As it encounters the signs, it may or may not know the meaning of each of them in terms of the instruction they represent with respect to way finding. The architecture is designed in such a way that if the sign is unknown, the robot explores its environment as a 'reflex' to find the next sign in the sequence from its memory. Once this next sign is discovered, the robot is able to denote a "meaning" ("turn right" or "turn left") to the previous sign, since this sign was supposed to lead the robot to the newly discovered sign. This meaning is then learnt for future reference. In this way, the robot continues to learn the meaning of new signs after deducing them. In following encounters with the same signs, it just executes the associated action based on its learning.

B. Overview of the system

The overall architecture proposed in this paper integrates the pre-exploration floor plan analysis into an organized neural structure (Fig. 2). It is based on a perception-action mechanism that constantly evolves because of the dynamic interaction between the robot and its environment [3].

The analysis of the floor plan in real time, represented by the box in the top half of Fig. 2, undertakes a thorough process permitting the robot to extract the relevant information for its integration into the system. It follows a process similar to the systems found in the literature [4]. It consists of an information segmentation process which identifies and separates different types of information; followed by structural analysis where the information is extracted (walls and navigation signs separately); and finally a semantic analysis where the information is extracted (walls and navigation signs

Figure 1. (1a.) Nao robot reading the map. (1b) Generated path by following the navigation signs. (1c) Real time view of the environment
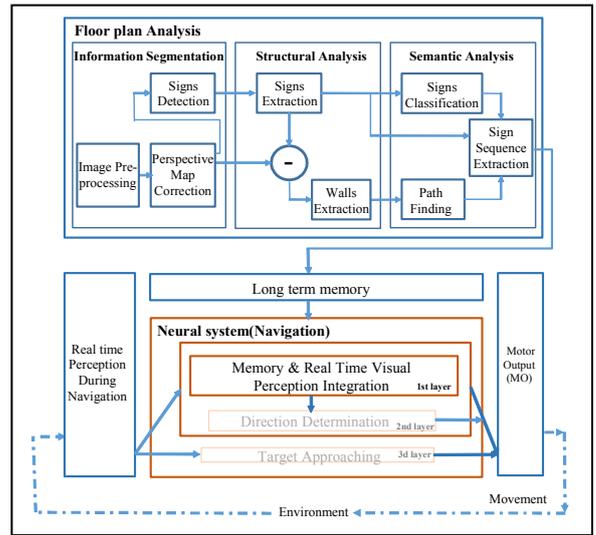

Figure 2. Document-based architecture

separately); and finally a semantic analysis allowing the extraction of the sign sequence based on the computation of the path and the information of the signs.

The neural system, shown in the lower half of Fig. 2, is composed of three different layers. The first layer of the system is in charge of merging the information coming from the long term memory and the real-time visual perception of the environment to navigate the robot towards its final destination. This layer is the major contribution of this paper and is explained in the next subsection. The other two layers, which are omitted in this paper for the sake of brevity in favor of detailing the more relevant contribution (information merging and document analysis processing), use a reflex mechanism that controls directly the robot's action based on the perceived information of the environment. Hence, the second layer determines the direction to be taken by the robot (left or right) by using the next expected sign from the pre-captured sequence, while the third layer directs an approach towards the sign by keeping it in the center of the robot's vision when the robot is far from it. More information about these two layers can be found in another paper by the same authors [5]. The whole system works in parallel and a 'competitive mechanism' allows to decide on the best behavior (among the layers) for controlling the robot according to the stimulus received. This is possible because the neural interconnection is done by either excitatory or inhibitory connections allowing or preventing the activation of neurons respectively.

C. Memory and Visual Perception Integration (first layer)

In order to reach its final destination, the robot needs to be able to use the information extracted from the floor plan and compare it to what it sees during real world exploration. Therefore, it has to be able to combine both sources of information and execute some actions accordingly. To that end, a system composed of artificial neural networks is implemented. The network comprises of interconnected artificial nodes known as "neurons" which compute their output values from one or more inputs received, akin to a biological neural network in the nervous system of an
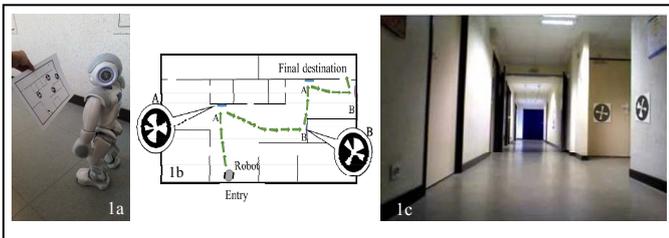
organism. These computational models have properties such as associative memorization, learning and parallel multi-information processing. Their interconnectivity is such that the whole system is capable of self-adapting with regard to its inputs and thus develop news behaviors.

Fig 3 shows the memory and visual perception integration layer as constructed with neural networks. It is composed of a number of interconnected groups of neurons. The first five groups correspond to the sensory perception and comprise of a number of neurons greater than the number of signs that can be recognized by the robot, each sign being represented by a single neuron. Thus, should the environment change with a different set of signs, the system will be able to represent them by using new neurons. The output direction group is composed of two neurons corresponding to the output motor actions (left and right movements respectively) and the reset group is composed of only one neuron. The flow of information through this unit and that represented in Fig. 3 is explained below.

There are two sources of information: the sign sequence extracted from the floor plan (long term memory) and the signs recognized by the robot in the environment during navigation (visual perception). Each one is computed in real time and stored in the "*sign sequenced group*" and the "*sign detection group*" respectively. Thereafter, a third neural group "*sign merged detector group*" combines both sets of perceptual information. Thus, when exploring the environment two scenarios are possible: the next expected sign in the sequence is detected, or not. In the former case, if the sign has already been associated to a particular movement, the robot executes the related movement i.e. turn left or right directly. On the other hand, if the sign has not been learnt previously, the robot uses the position of the next expected sign with respect to the current one to associate a direction instruction to this current sign. Therefore, it is necessary to keep the current sign in memory for a short while the next one is being looked for. This information is memorized in the "*short term memory group*" and is deleted by the "*reset group*" once the next sign is found and the movement is performed. The generation of the said association takes place in the "*winner-takes-all (WTA) group*" which links the information of the recognized sign and the reflex movement performed in the second layer (Direction Determination (DD)). This group also allows the learning of such associations by activating and sending a reinforcement signal. Finally the information of the action associated to the recognized sign is stored in the "*output direction group*" and sent to the motor output.
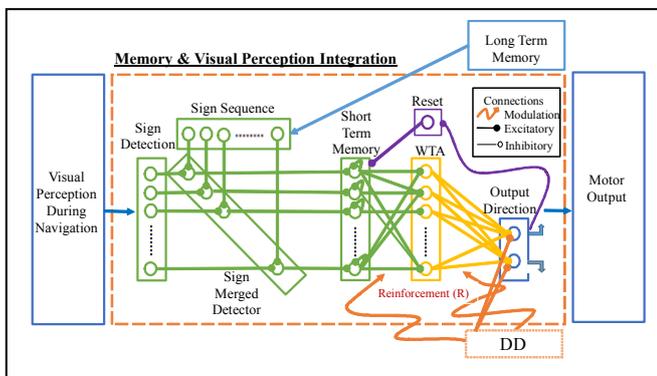
III. FLOOR PLAN ANALYSIS

When working in real time situations, reading and analyzing a document from an image acquired from a mobile device such that of a robot is a challenging task. It not only has to be accurate but it also has to be fast so that the robot can react accordingly. Image acquisition of documents under constrained and dynamic conditions may cause not only degradation but also distortions on the captured image itself due to unpredictable orientation, distance, lighting conditions etc. Different scenarios might be possible. For instance in the case of this work, since there is no use of a planar support when acquiring the image, the map might not be completely or at all within the image frame. It might also be too small or suffer from perspective distortion if the image captured is not orthogonal to the plane of the camera (Fig.4). All these circumstances make it hard to properly identify and retrieve the content of the map. Therefore, a set of processes after image acquisition has been implemented (top half of Fig. 2).

Usually maps and floor plans contain building elements like walls, doors, windows, etc. and textual information such as the name of the room or the area. The focus of this work is to read a modest floor plan in the simplest way in order to prove the concept of robot navigation as explained above. Therefore, the representation of the doors in the design of the floor plans is done by blank gaps and only the walls (external and internal) as well as the navigation signs are here considered (Fig 4).

A. Information segmentation

As the outcome of the segmentation process directly affects the performance of further processing, it is important to find a suitable *image pre-processing* algorithm that allows removal of any noise and help to simplify the representation of the image for better analysis.

A global thresholding has been applied to the images in this work, as it has proven to be the simplest of image segmentation methods while providing a fast and convenient way to perform a good result when working in real time situations [6]. However, non-uniform illumination of the document may cause inaccuracies or complete failure of segmentation. In order to overcome these effects, a black top-hat morphological operator with a 7*7 square mask is first applied on the gray-scale image. This process consists of removing noise and small objects from the image that are not relevant to the analysis of the image. Hence, it improves the clarity of the image outcome (see Fig. 5) and consequently, it makes possible the identification of relevant regions (map) and objects of interest (navigation signs and walls) in the image.

The identification and then extraction of the map from the image is made possible using analysis of the connected components. This process together with the application of

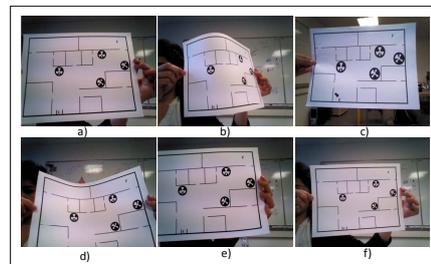Figure 3. Combination between document and visual information

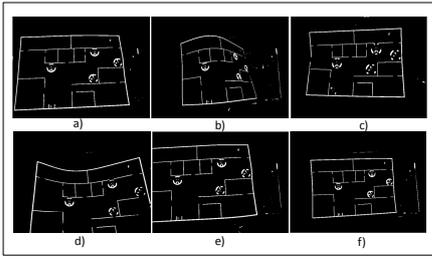Figure 4. Different scenarios of image acquisition

Figure 5. Binary Image by thresholding

some filters based on the size of the connected regions, permits rejection of images likely to contain irrelevant components outside the convex hull (Fig. 5b) or with missing information like an incomplete map boundary (i.e. user's fingers holding the map as in Fig. 5c or map outside the camera frame as in Fig. 5d and 5e). Hence, only the images containing the complete and visible map are selected (Fig. 5a and 5f).

Several methods have been proposed for image *perspective correction*. Some of them aim to calculate the horizontal and vertical vanishing point based on the horizontal lines of the text or the stroke boundaries [7] [8]. Others, instead, aim to find the corners of the quadrilateral formed by the boundary between the background and the document [9] [10]. The latter method is more suitable for the type of document used in this work. Therefore, the method proposed in this work consists of first creating a bounding box around the image (top half of Fig. 6), then retrieving the corners coordinates (X,Y) of the bounding box (rectangle) formed to use them as reference points to find the coordinates (x, y) of the corners of the map contour. By calculating the minimum distance between the coordinates (X, Y) of each corner of the bounding box and each vector point of the map contour, the closest vector point to each corner of the bounding box is found. This gives the corners (x, y) of the map contour. The perspective correction is then performed by using a 3x3 transformation matrix calculated with the coordinates of quadrilateral vertices of the map and those of the destination image (bottom half of Fig. 6).

The navigation signs are represented in this work by circle-shape symbols containing specific patterns that allow them to be differentiated from one another. Hence, in order to distinguish them from the floor plan, *sign detection* is achieved using a feature based approach by using the Hough circles transform [11] (white circles around the signs in bottom half of Fig.6).

B. Structural Analysis

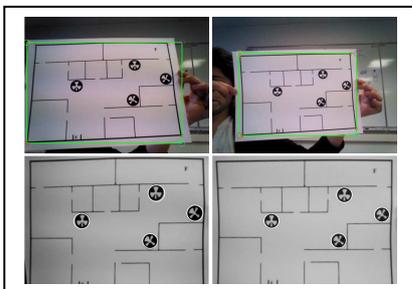The information obtained in the previous step is used in

Figure 6. (Top) Selected Images (5a and 5f) for map extraction. (Bottom) Navigation signs detection after Perspective correction

order to extract each sign from the floor plan. By considering the parameters referring to the center position and radius of the circles (denoting signs) obtained previously (y_{center}, x_{center}, r), it is possible to subtract the information within each circle from the floor plan (Fig. 7a) and copy each of the signs into a new image. As a result, a 'walls image' composed of only internal and external walls is produced (Fig. 7b), as well as a set of navigation sign images each consisting of a unique sign.

C. Semantic Analysis

In order to compute the correct sequence of signs, two different process are simultaneously applied to the resulting images from the last step. In the first one, the robot recognizes the signs images. There are two notable methods available for the recognition of signs in maps or paper-based documents using image processing: keypoints based [12] [13] or template matching based [14]. The former aims to extract some points of interest (generally the corners) and describe them using some radiometric features. The latter aims to find small parts of an image that match a template image from existing database. This latter simpler approach was used in this work since a database of the signs images was already available and could easily be recognized by the robot in the environment.

Simultaneously, using the other process, the robot is able to compute a navigation path to reach the final destination in an environment with obstacles (walls on the map). The robot uses a neural network implementation of the resistive grid technique [15]. The image of the walls is used here as a resistive grid where each pixel in the image is represented by a neuron and is connected to its closest neighbors. Each neuron has an activation value which is calculated based on the distance between it and the neuron representing the final destination. This latter having the highest value "1" and those representing the walls the minimum "0". The path is computed by continuously searching among neighboring neurons for the one with the highest activation value. This continues until the final destination is found (Fig.7c).

Finally the sequence of the navigation signs is calculated by using the position of the recognized signs with respect to the generated path, from the starting point to the final destination. To achieve this, certain regular intervals of distance, 'checkpoints' (in terms of x-y coordinates), are first demarcated along the path. Beginning at the starting point of the extracted path, for each checkpoint, the distance to all visible signs is calculated. The sign located at the shortest distance for each checkpoint is stored in a special array. This is done for all checkpoints until the final destination. As a result, the array comprises of all the navigation signs visible along, and closest to the path at each checkpoint. It is then possible to finally extract the right sequence of signs that the robot will encounter on its way (Fig.7d).
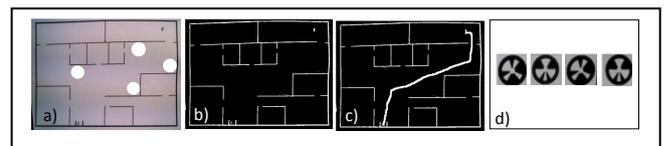

Figure 7. Navigation signs and walls segmentation

IV. EXPERIMENTS IN REAL ENVIRONMENT

The complete neural architecture was built, implemented in the humanoid robot NAO and tested in a 2.25m x 4.5m space, configured such that it represented a simplified layout of a floor of a building composed of a room delimited by walls (here tables) and a corridor between them. The intended navigation path was free of dynamic obstacles and only the walls were considered as the static obstacles to be avoided. The NAO robot can detect a sign using its camera (640*480 pixels) if it is within a range of approximately 0.2m to 1.5m away from it. Therefore, the navigation signs were placed in the environment within this range so as to always be readable.

Before exploration, the robot was shown a floor plan of the test environment so that it could process the image and compute the best way to accomplish the task. By following the steps of floor plan analysis explained in section III, the robot was able to compute the path leading towards the final destination and thus extract the sign sequence :{ A, B, A, B} as shown in Fig.8. When the map was not placed correctly within the frame of the camera, the robot was able to warn the user to move it. Otherwise, when it was correctly positioned, the average computation time to extract the information was of around 5s.

Thereafter, the robot was placed at the entrance of the test environment from which it could distinctly see the first expected sign, *A*. While navigating, the following behaviors emerged allowing the robot to successfully reach its goal:
• When a sign was detected, it was able to compare it to the corresponding sign in the extracted sequence from the memory;
• When the comparison resulted in a positive result i.e. the detected sign was indeed the expected sign, and the meaning of the sign was yet unknown, it was able to perform reflex movements by rotating in one place to search for the next sign, figure out the associated direction and then learn it;
• When the comparison gave a negative result i.e. the detected sign did not match the expected sign, the robot ignored the detected sign and continued reflex movements to locate the correct sign;
• When the expected sign was not in its visual field, it was able to perform a reflex behavior to search for it;
• When a sign was faraway, it was able to get closer;
• When a sign appeared again, it was able to recall the learning and perform the associated movement.
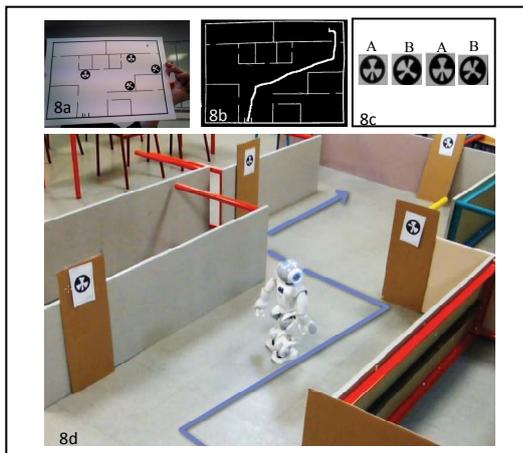

Figure 8. (8a.)Map of the building to be explored (8b) Path found (8c) sign sequence. (8d) Navigation of the robot towards the goal

V. CONCLUSIONS

In this paper, a robot uses a chain of image processing techniques for the analysis of a floor plan in real time. Relevant information is extracted and later compared with the visual information perceived during real world navigation. Such an approach at the same time top-down and bottom-up is merged into a neural system allowing the robot to reach its goal faster while still adapting itself to an environment prone to change. Different behaviors emerging from the interaction between the robot and the environment prove the advantages of a neural approach emulating human behavior.

Experimental results have shown that the use of a floor plan prior to navigation can facilitate and hasten the task of reaching a specific destination inside a building for a mobile robot.

An eventual scenario could be one where the expected signs are no longer available or visible inside the building. In that case, the robot will be able to generate an updated map according to the information assimilated from the environment and still achieve its final destination.

REFERENCES

[1] R. Siegwart, I.R. Nourbakhsh, and D. Scaramuzza. "Introduction to autonomous mobile robots". MIT press, 2011.

[2] F. Qureshi, D. Terzopoulos and R. Gillett, "The cognitive controller: a hybrid, deliberative/reactive control architecture for autonomous robots". Innovations in Applied Artificial Intelligence, 2004.

[3] P. Gaussier and S. Zrehen, "PerAc: A neural architecture to control artificial animals",Robotics and Autonomous Systems, vol. 16, no. 2-4, pp. 291-320, 1995.

[4] S. Ahmed, M. Liwicki, M. Weber, and A. Dengel. "Improved automatic analysis of architectural floor plans". In Document Analysis and Recognition (ICDAR), IEEE, pp. 864-869. 2011.

[5] D.M. Rojas-Castro, A. Revel, M. Ménard. "A robust neural robot navigation using a combinationof deliberative and reactive control architectures". European symposium on Artificial Neural Networks (ESANN), 2015.

[6] K.K. Singh, and A. Singh. "A study of image segmentation algorithms for different types of images" Int. J. Comput. Sci. Issues vol.7, pp.414-7 2010.

[7] P. Clark, and M. Mirmehdi. "Rectifying perspective views of text in 3D scenes using vanishing points". Pattern Recognition vol. 36, no 11, pp. 2673-2686 ,2003.

[8] S. Lu, B.M. Chen, and C.C. Ko. "Perspective rectification of document images using fuzzy set and morphological operations". Image and Vision Computing, vol. 23, no 5, p. 541-553,2005.

[9] P. Clark and M. Mirmehdi, "Recognising text in real scenes". International Journal on Document Analysis and Recognition, vol. 4, no. 4, pp. 243-257, 2002.

[10] J.Rodríguez-Piñeiro et al. "A new method for perspective correction of document images." IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, pp. 787410-787410, 2011.

[11] D. Ballard, "Generalizing the Hough transform to detect arbitrary shapes". Pattern Recognition, vol. 13, no. 2, pp. 111-122, 1981.

[12] M. Rusinol and J. Llados. "Symbol Spotting in Digital Libraries: Focused Retrieval over Graphic rich Document Collections". London: Springer, 2010.

[13] M. Rusinol, D. Karatzas, and J. Llados. "Spotting graphical symbols in camera-acquired documents in real time". In GREC proceedings, 2013.

[14] J. Weber and S. Tabbone. "Symbol spotting for technical documents: Ancient template matching approach". In Pattern Recognition (ICPR), 21st International Conference, pp. 669-672, Nov 2012.

[15] G.Bugmann, J.G. Taylor, and M. Denham. "Route finding by neural nets". Neural networks,pp. 217-230,1995.