Combining structural and statistical strategies for unsupervised wall detection in floor plans

Lluís-Pere de las Heras, Ernest Valveny, Gemma Sanchez
Computer Vision Center - Dept. Ciències de la Computació
Universitat Autònoma de Barcelona
Barcelona, Catalonia, Spain
Email: {lpheras,ernest,gemma}@cvc.uab.cat

***Abstract*—This paper presents an evolution of the first unsupervised wall segmentation method in floor plans, that was presented by the authors in [1]. This first approach, contrarily to the existing ones, is able to segment walls independently to their notation and without the need of any pre-annotated data to learn their visual appearance. Despite the good performance of the first approach, some specific cases, such as curved shaped walls, were not correctly segmented since they do not agree the strict structural assumptions that guide the whole methodology in order to be able to learn, in an unsupervised way, the structure of a wall. In this paper, we refine this strategy by dividing the process in two steps. In a first step, potential wall segments are extracted unsupervisedly using a modification of [1], by restricting even more the areas considered as walls in a first moment. In a second step, these segments are used to learn and spot lost instances based on a modified version of [2], also presented by the authors. The presented combined method have been tested on 4 datasets with different notations and compared with the state-of-the-art applyed on the same datasets. The results show its adaptability to different wall notations and shapes, significantly outperforming the original approach.**

I. Introduction

Wall detection is a crucial step in floor plan interpretation because walls globally define the structure of buildings. Nevertheless, due to the lack of a standard graphical notation in floor plan's modeling, there is a big variability on how these elements are lineally drawn; they can be represented by thick black lines, parallel lines, hatched textures, etc.. This issue has not been solved by classical approaches, that are very adhoc for a small set of notations and therefore, useless for the rest.

Several approaches have been proposed for wall segmentation in the literature focusing on specific notations. In [3],[4],[5],[6], different strategies are followed for the sake of segmenting walls modeled by thick black lines. Both, [3], [4], apply morphological filtering to thin-thick line separation. Contrarily, [5], [7] use Hough Transform over the vectorized image to detect parallel lines with black texture in between. While [6] finds the potential walls by studying the polylines generated after vectorizing only thick lines in the image.

All of these floor plan recognition systems need to redefine the wall segmentation step when dealing with images of different notations, as the ones represented by simple parallel lines [8], or a hatched texture [9]. This fact not only provokes that the floor plan recognition problem can not be considered as solved yet, but also that the different approaches are not even comparable since most of the approaches are significantly oriented to their own notations. With the aim of finding a solution to this general extended problem when dealing with this sort of documents, our group has put the effort in the recent years on creating an effective wall detector able to segment walls in all the floor plans, independently of their graphic notation.

The first attempt was inspired by the appearance-based state-of-the-art strategies in Computer Vision for object detection in real scenes. As a result, in [9] a bag of visual patches able to learn the visual appearance of walls from a labeled collection of floor plans was presented as the first approach to segment walls in multiple collections of real documents. This system was refined and its performance enhanced in [2], showing a great adaptability to noise and notations. Even though the approach demonstrated its suitability in real cases when dealing with controlled sets of floor plans and notations, the need of generating the ground-truth each time a new collection was added to the system, pushed us to rethink the whole strategy.

The second attempt was to tackle the problem under a structural point of view. The latent idea is to be able to drive the detection of the potential elements belonging to walls by general structural properties of buildings and thus, without the need of any learning step for each notation. In [1], the authors present an unsupervised approach driven by six structural rules of general properties of buildings, that combined fuzzily are able to segment walls in different collections of floor plans with successfully close results to the supervised approach. With all, the method still has some restrictions inherent to the structural properties, e.g "walls are usually straight" or "longer than thicker".

In this paper we close the circle by proposing a new method as a result of the combination of both strategies, the structural one and the statistical one, that is able to detect walls independently of notation and structure, and without the need of any previously labeled data. Firstly, with a modified version of the approach described in [1] we return a wall image with an initial segmentation of the floor plan. Then, a modified version of the supervised approach presented in [2], learns the appearance of walls and redefines the initial segmentation. The proposed method has been tested on the same datasets and Internet images than [1], demonstrating a significant improvement in the detection on floor plans with curved walls, and a better performance in the overall datasets.

In Section II we explain the method, revisiting the two

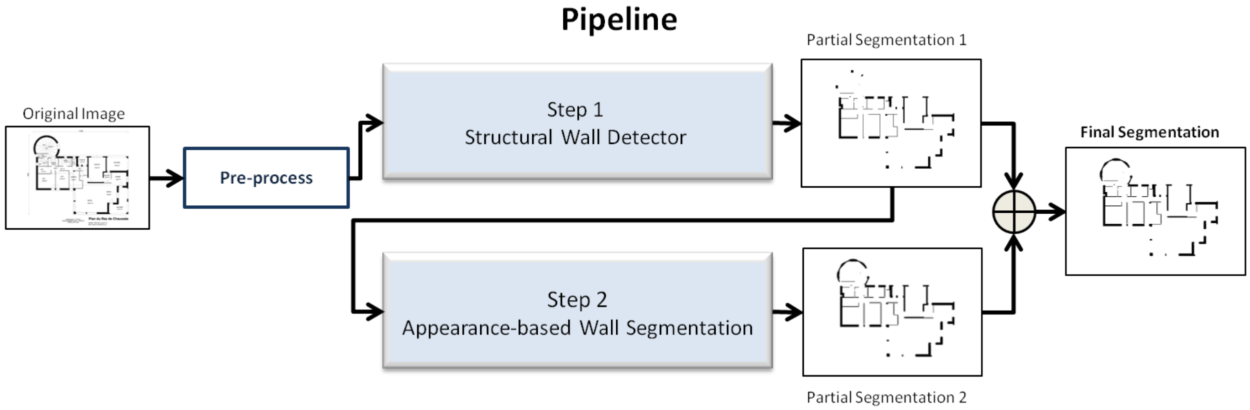

Fig. 1: Pipeline of the method

approaches combined. Section III is devoted to present the experimental evaluation. Finally, in Section IV we conclude the paper.

II. METHODOLOGY

The detection is driven by these 6 general assumptions of walls in building drawings:

1) Walls are drawn by parallel lines.
2) They appear in orthogonal directions.
3) Walls are rectangularly shaped, usually longer than thicker.
4) They define the structure of the building; appearing naturally distributed in the plan.
5) Different thickness are used to model internal and external walls.
6) The walls in a document are filled by the same pattern (hatched, tiled, solid, etc.).

These assumptions cannot be seen as a complete pack of unbreakable statements for all existing real floor plans. For example, there are floor plans with diagonal or curved walls, buildings with the same thickness for interior and exterior walls, etc. Nevertheless, a relaxed combination of them enhances the flexibility of the system, leading to a good final segmentation independently of the building or document complexity.

Our method, whose pipeline is shown in Figure 1, is the result of combining the recent wall detector methods [1] and [2], driven by the wall's general knowledge postulated above. It can be separated in two steps: a structural-based detection and an appearance-based detection. In the first step, we extract high confident wall segments according to their structural properties. Then, in the appearance-based step based on a modification of [2], these segments are used to learn their visual appearance and so, refine the final segmentation. In this section we explain the complete methodology, overviewing both original methods involved by specially focusing on those aspects that have been modified to accomplish our final solution.

A. Step 1: Structural Detection

The first stage of the method is devoted to detect wall segments by their structural properties, similarly to [1]. The detection starts by detecting elements formed by parallel lines, according to *assumption 1*. The confidence of the resulting wall segments is determined by their agreement with the *assumptions 3*, *4* and *5*. This process is divided into *Preprocessing*, *Black-wall detection*, *Wall-segment candidates* and *Confident wall segmentation*.

Preprocessing

The images are binarized and the textual information is filtered out using [10]. Possible deviations in the floor plan modelling are corrected by an adaptation of [11]; a method for handwritten deskewing. This is done to facilitate the detection of parallel lines in the orthogonal directions, as *assumption 2* asserts. Even though the segmentation strategy is scale-invariant, for efficiency issues the images with resolutions higher than 4000$\times$4000 pixels are down-scaled.

Black-wall detection

Some old floor plans model walls by black thick lines as the ones shown in Figure 3a. Since the preliminary detection is based on finding parallel lines, an automatic adhoc preprocessing has to detect and transform these sort of documents to a more suitable input. Firstly, horizontal and vertical runs of foreground pixels are quantized in a histogram. Documents with thick walls present sparser histograms with far more outliers in higher positions than the rest. Then, they are easily detected by thresholding the sigma of the Gaussian mixture fitted to the histogram by EM. Images with black thick walls are replaced by their edge image, obtained using the Canny edge detector.

Wall-segment candidates

Wall-segment candidates are seek in a first step according to assumption 1; seeking parallel lines in the document. The lines in the document are encountered by only considering those runs of foreground pixels of a certain minimum length. They are quested in multiple orientations of the image. The

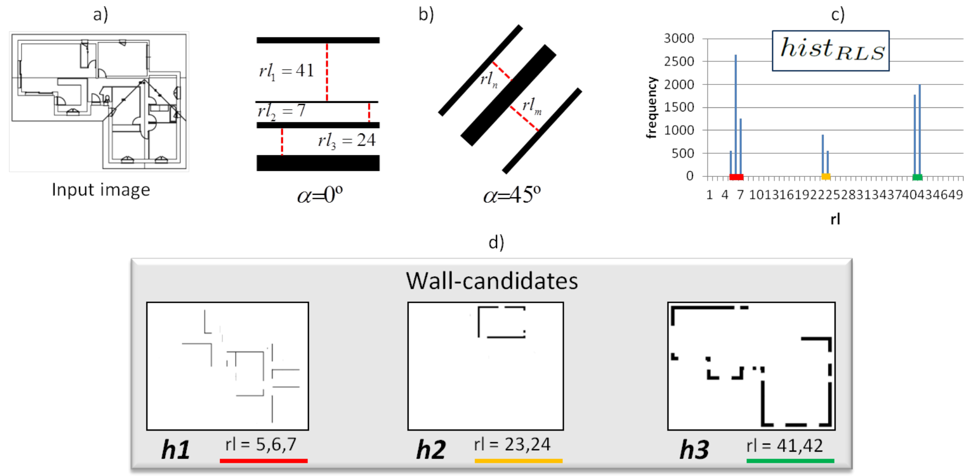

Fig. 2: Wall candidate generation. The input image is shown in a). In b), the run extraction process at two different orientations α is zoomed. This runs are quantized in the histogram shown in c), generating three colored clusters that belong to common parallel line thickness in the input image. For each cluster, the parallel lines of their corresponding thickness are retrieved in three different candidates h.

distances between lines are quantized in a histogram, where bins with higher frequencies stand for common parallel line distances. The histogram is smoothed in order to filter out irrelevant information. Then, for each non zero bin, a candidate segmentation image is generated by retrieving the areas according the corresponding distances between parallel lines. In other words, each segmentation image contains wall segments of the same thickness. This process is detailed in Figure 2.

Confident wall segmentation

To extract the segments with the highest degree of confidence, we firstly rank the wall image candidates before spreading them into the final segmentation. This is done oppositely from the original approach, where all possible wall candidates were spread to posteriorly rank them regarding they confidence. The reason is that the wall's texture will be learned from these segments and used to recover lost instances that do not agree with the structural *assumptions*. Thus, at this point the precision on wall segmentation has to be maximized to enhance the posterior visual learning step. It is also worth to point out that most of the confident segmentations tend to rely on exterior and interior walls, as it is assumed in *4*.

Let H be the set of wall image candidates h, where the segment thickness $thick(h_i) \neq thick(h_j)$ for all $i \neq j$. Agreeing with *assumption 4*, walls are elements that appear repeatedly in floor plans. The score Ncc accounts the number of segments in the segmentation image:

$$\mathrm{Ncc}_{h_i} = \#\mathrm{CC}(h_i). \tag{1}$$

Likewise, according to the *assumption 3*, walls are longer than wider and then, segmentations with longer segments are most likely to be a correct. Thus:

$$\mathrm{AR}_{h_i} = \overline{\mathrm{long}(\mathrm{CC}_j(h_i))/\mathrm{width}(\mathrm{CC}_j(h_i)), \forall j | CC_j \in h_i} \tag{2}$$

where AR_{h_i} bears the aspect ratio of connected components (CC).

Finally, also according to *assumption 4*, **DiffD** enforces segmentations with similar black pixel distributions to the input image:

$$\mathrm{DiffD}_{h_i} = \sum_{n=1}^{r} \sum_{m=1}^{r} p_{nm} - p_{nm}^{h_i}. \tag{3}$$

where p_{nm} and $p_{nm}^{h_i}$ are the percentage of the black pixels in the nm^{th} region r of the original image and h_i respectively.

The final rank is calculated by sorting in a descend means the candidates by their global score, which is calculated as:

$$W(h_i) = \mathrm{Ncc}_{h_i} + \mathrm{AR}_{h_i} + \mathrm{DiffD}_{h_i}. \tag{4}$$

Once the ranking of the segmentation candidates is done, the top n are combined into the final segmentation image. The number of selected images n depends on a experimentally calculated boundary over the confidence score. This boundary is thought to not only enforce final segmentations with more than only one candidate (assumption 5), but also discard completely impossible segmentations.

B. Step 2: Appearance-based Detection

Up to now we have an image of wall segments that we call segment-image for clarity. The aim in this step is to learn the visual appearance of walls from it, and refine the final segmentation by overtaking strict statements, such as walls are straight elements. The process here is similar to the one from [2], but reconsidering the learning step; it is done from a single image instead from a pre-anotated corpus of images. In this section we explain the learning procedure and summarize the classification.

Learning

The original image is split into squared equal-sized and overlapped patches. This procedure is repeated for the image rotations 45°, 90° and 135° with two purposes: to get

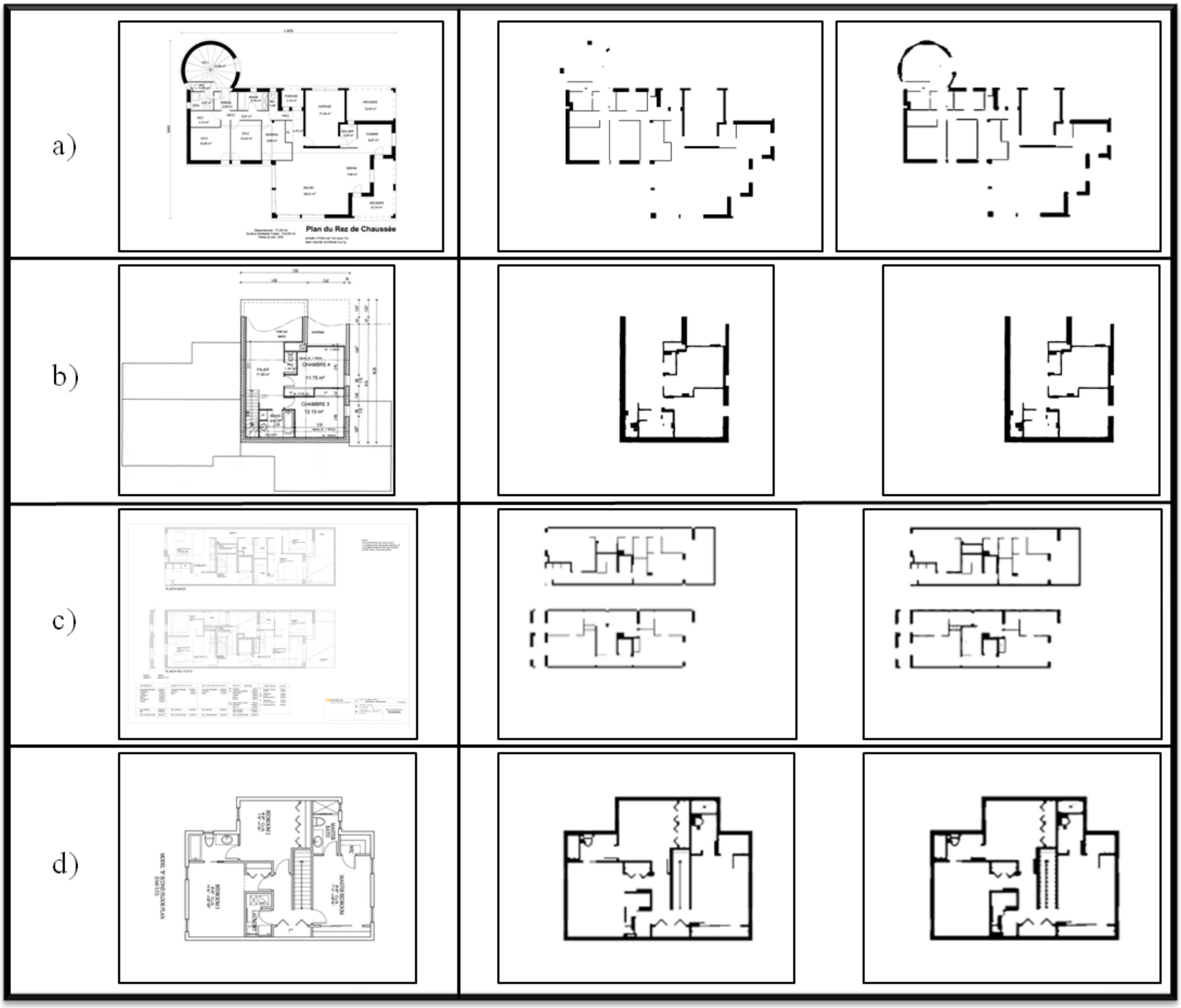

Fig. 3: Wall segmentation examples. Regarding the dataset collection a) *Black*, b) *Textured*, c) *Textured2* and d) *Parallel*. From left to right: the original image, the preliminary segmentation after Step 1 and the final segmentation obtained after Step2.

more learning instances and to achieve rotation-invariantness. Patches falling into segmented regions in the segment-image are labeled as positive examples $c = \{Wall\}$, the rest as negative $c = \{Background\}$ meanwhile completely white patches are filtered out. The image descriptor BSM [12] is used to describe them. Then, a subset of K patch-descriptors pd, that contains the same number of positive and negative instances, are clustered into a dictionary of visual words using a fast version of Kmeans [13].

Once the dictionary is created, a probability to each word w in the dictionary is calculated regarding its representative class $C = \{\mathrm{Wall}, \mathrm{Background}\}$. Every patch-descriptor pd_i, that has already a label belonging to one of the two classes calculated previously, is assigned to its closest word in the dictionary w_j. Then, the conditional probability of a word to belong to each one of the classes is given by:

$$p(c_i|w_j) = \frac{\#(pd_{w_j}, c_i)}{\#pd_{w_j}}, \forall i, j. \tag{5}$$

Where $\#(pd_{w_j}, c_i)$ states for the number of patch descriptors with the label c_i assigned to codeword w_j, and $\#pd_{w_j}$ is the total number of patch-descriptors assigned to w_j.

Recognition

Every patch-descriptor from the overlapped grid on the input image inherits the class probabilities of its nearest word in the dictionary. This classification is performed by the 1-NN hard assignment on the Euclidean space. Lastly, the final pixel categorization depends on the multiple patches that fell on it. The *Mean Rule* on the theoretical framework of combining multiple classifiers is adapted to calculate the final segmentation for every pixel px in the image:

$$class(px) = \arg\max_i \Big(mean(P(c_i|pd)) \Big), \forall pd \mid px \in pd. \tag{6}$$

III. Experiments

In this section we explain the experimental evaluation performed. We firstly overview the datasets and the evaluation protocols, which are exactly the same as in [1]. And then, we present the results quantitatively and qualitatively.

A. Dataset

The dataset used to evaluate our method is the same as in [1]. It contains 4 different datasets of real floor plans together with some images randomly picked from the Internet[1]. Let us summarize the four datasets

- **Black**. This dataset contains 90 images of real floor plans and has been used for evaluation of wall detection in [9], [2] and [1]. Walls are black-thick lines of different thickness whether they are interior or exterior. An example image of this dataset is shown in Figure 3a.
- The **Textured** collection contains 10 noisy real floor plans. Here, walls are modeled with different texture depending whether the are interior and exterior, as it can be seen in 3b.
- **Textured2**. This dataset contains 18 high resolution real floor plans. Walls contain multiple thickness, for interior, exterior and main walls textured with a hatched pattern. 3c shows an instance of this collection.
- **Parallel** contains 4 real floor plans. Walls are modeled by simple parallel lines, without any texture inbetween. An image of this set is shown in 3d.

B. Evaluation protocol

The evaluation protocol is the same used in the last wall detection works [9], [2] and [1]: the Jaccard Index (JI). As in [1], in addition to JI the global recall is also taken into account for the experimental evaluation. The reason is that, if we consider this method as a single step from a complete system of floor plan interpretation, higher recall results are preferred, as false positives are easily cleared in post-processing than detecting lost instances. The JI and Recall are calculated respectively as:

$$JI = \frac{TruePos}{TruePos + FalsePos + FalseNeg},$$

$$Recall = \frac{TruePos}{TruePos + FalseNeg}.$$

[1] https://www.google.es/imghp?q=floor%20plan

C. Experimental Results

Our method is inherently affected by the same parameters than the original approaches. In the first step of our method, the parameter values considered in [1] are also adopted here for comparison purposes. Hence, rl^b_{min}, which indicates the minimum length of the runs on black pixels to be considered as lines, is set to 10 pixels. The rotation angle interval α that defines the orientation where the lines are seek is defined to $15°$. The boundary value of σ^{thw} to discriminate floor plans with thick black walls is 25. And finally, the number of equal-sized squared divisions r to calculate the difference on the pixel distribution (DiffD) is 9.

On the other hand, the parameters in the second step have been restudied and recalculated experimentally since the learning origin is completely different from [2]. The parameters that affect the behavior of our method are three inherited from the original approach: the size of the patches PS, the distance between patches ϕ^{ov}, and the dictionary size DS; and a fourth one generated by the new learning framework, that accounts the amount of patches used for creating the vocabulary Spd. Regarding PS, only proportional values to the highest wall thickness in the segment-image have been tested, adopting finally 0.5 times the size of the thickest segment. ϕ^{ov} measures the distance between the centers of the patches or, in other words, the grid overlapping. Here, several proportional values to patch-size have been tested being $1/2PS$ the value which lead to the best performance. In terms of the dictionary size, smaller dictionaries proved to generalize better. Thus, just 300 words are enough to learn the wall texture. Finally, the number of patch-descriptors Spd used in the learning needs to be defined. The experiments have shown that the more learning data, the better results. Nevertheless, we have detected a point where the classifier saturates. This saturation point is up to 75.000 patch-descriptors, half of them labeled as *Wall* and the rest as *Background*.

Table 1 shows the quantitative results obtained by our new approach compared with the most recent state-of-the-art working on the same set of the images. From the results we observe that the proposed unsupervised approach performs better than the original unsupervised approach [1]. The performance is not only better on the average recall but also on JI terms in two of the datasets. On the other hand, even the higher recall, our new method still behaves slightly worse in JI terms than both supervised strategies, [9] and [2]. Finally, it is worth to say that [3] is notation-oriented approach specifically thought for dataset *Black*, which makes it useless in the rest of images.

Wall segmentation on an example from each dataset is shown in figure 3. Moreover, three challenging images extracted from internet are shown in figure 4. These images were already used to justify the adaptability of [1]. Both figures illustrate the adaptability of our new method independently on different wall notations, shapes and noise.

IV. Conclusion

In this paper we have presented an unsupervised wall segmentation method based on the combination of two recent approaches. On top of the original segmentation method [1], a Bag-of-Patches step [2] has been used to learn the visual appearance of walls and to refine the final segmentation. Thus,

TABLE I: Wall segmentation results

	#images	[3]		[9]		[2]		[1]		**our new approach**	
		JI	**Rec.**	**JI**	**Rec.**	**JI**	**Rec.**	**JI**	**Rec.**	**JI**	**Rec.**
Dt. Black	90	0.90	0.92	0.97	0.99	0.97	0.99	0.93	0.97	0.95	0.99
Dt. Textured	10	–	–	0.83	0.98	0.86	0.99	0.82	0.97	0.82	0.98
Dt. Textured2	18	–	–	0.81	1	0.82	1	0.77	0.91	0.79	0.96
Dt. Parallel	4	–	–	0.70	0.84	0.71	0.86	0.66	0.98	0.67	1
Average per Dataset		–	–	**0.83**	**0.95**	**0.84**	**0.96**	**0.80**	**0.96**	**0.80**	**0.98**

(a) (b) (c) (d) (e) (f)

Fig. 4: Qualitative results for images extracted from the Internet. In (b), (d) and (f), the segmented walls are shown from their corresponding original images (a), (c) and (e).

the imposed structural restrictions are relaxed and the elements usually lost by the original method, such as curved walls or beams, are correctly segmented here. We have compared its performance with the most recent wall segmentation strategies in four different floor plan datasets, and some other images downloaded from the Internet. The results show its great adaptability to different image notations and resolutions and without the need of any labeled data to learn the wall notation each time.

ACKNOWLEDGEMENT

This work has been partially supported by the Spanish project TIN2009-14633-C03-03 and by the research grant of the Universitat Autonòma de Barcelona (471-02-1/2010).

REFERENCES

[1] L.-P. de las Heras, D. Fernéz, E. Valveny, J. Lladós, and G. Sánchez, "Unsupervised wall detector in architectural floor plan," in *Proceedings of the 12th International Conference on Document Analysis and Recognition*, 2013, pp. in–print.

[2] L.-P. las Heras, J. Mas, G. Snchez, and E. Valveny, "Notation-invariant patch-based wall detector in architectural floor plans," in *Graphics Recognition. New Trends and Challenges*, ser. Lecture Notes in Computer Science, 2013, vol. 7423, pp. 79–88.

[3] S. Ahmed, M. Liwicki, M. Weber, and A. Dengel, "Improved automatic analysis of architectural floor plans," in *Proceedings of the 11th International Conference on Document Analysis and Recognition*, 2011.

[4] P. Dosch, K. Tombre, C. Ah-Soon, and G. Masini, "A complete system for the analysis of architectural drawings," *International Journal on Document Analysis and Recognition*, vol. 3, pp. 102–116, 2000.

[5] S. Macé, H. Locteau, E. Valveny, and S. Tabbone, "A system to detect rooms in architectural floor plan images," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 2010, pp. 167–174.

[6] Z. Shi and V. Govindaraju, "Line separation for complex document images using fuzzy runlength," in *Document Image Analysis for Libraries, 2004. Proceedings. First International Workshop on*, 2004, pp. 306–312.

[7] L.-P. de las Heras and G. Sanchez, "And-or graph grammar for architectural floorplan representation, learning and recognition. a semantic, structural and hierarchical model," in *Proceedings of the 5th Iberian Conference on Pattern Recognition and Image Analysis*, vol. 6669, 2011, pp. 17–24.

[8] T. Lu, H. Yang, R. Yang, and S. Cai, "Automatic analysis and integration of architectural drawings," *International Journal on Document Analysis and Recognition*, vol. 9, pp. 31–47, 2007.

[9] L.-P. de las Heras, J. Mas, G. Sánchez, and E. Valveny, "Wall patch-based segmentation in architectural floorplans," in *Proceedings of the 11th International Conference on Document Analysis and Recognition*, 2011, pp. 1270–1274.

[10] K. Tombre, S. Tabbone, L. Pélissier, B. Lamiroy, and P. Dosch, "Text/graphics separation revisited," in *Document Analysis Systems V*, ser. Lecture Notes in Computer Science, 2002, vol. 2423, pp. 615–620.

[11] N. Ouwayed and A. Belaid, "A general approach for multi-oriented text line extraction of handwritten document," *International Journal on Document Analysis and Recognition*, vol. 14, no. 4, Sep. 2011.

[12] S. Escalera, A. Fornes, O. Pujol, A. Escudero, and P. Radeva, "Circular blurred shape model for symbol spotting in documents," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 2005–2008.

[13] C. Elkan, "Using the triangle inequality to accelerate k-means," in *the Twentieth International Conference on Machine Learning*, 2003, pp. 147–153.