Perceptual Retrieval of Architectural Floor Plan Images

Lluís-Pere de las Heras, David Fernández, Alicia Fornés, Ernest Valveny, Gemma Sànchez and Josep Lladós
Computer Vision Center - Dept. de Ciències de la Computació
Universitat Autònoma de Barcelona
Barcelona, Catalonia, Spain
Email: {lpheras,dfernandez,afornes,ernest,gemma,josep}@cvc.uab.cat

***Abstract*—This paper proposes a runlength histogram signature as a percetual descriptor of architectural plans in a retrieval scenario. The style of an architectural drawing is characterized by the perception of lines, shapes and texture. Such visual stimuli are the basis for defining semantic concepts as space properties, symmetry, density, etc. We propose runlength histograms extracted in vertical, horizontal and diagonal directions as a characterization of line and space properties in floorplans, so it can be roughly associated to a description of walls and room structure. A retrieval application illustrates the performance of the proposed approach, where given a plan as a query, similar ones are obtained from a database. A ground truth based on human observation has been constructed to validate the hypothesis. Preliminary results show the interest of the proposed approach and opens a challenging research line in graphics recognition.**

I. Introduction

Aesthetics is a branch of philosophy devoted to beauty, scientifically defined as the study of sensory or sensori-emotional values, sometimes called judgements of sentiment and taste. Visual aesthetics is an emerging topic in computer vision [1], [2]. The general idea is to use machine learning techniques to score the aesthetics of images in terms of visual cues as color, shape and texture, given a subjective human characterization. In document analysis, and in particular in graphics recognition, to the best of our knowledge the concept of aesthetics has not been developed. Some promising approaches have been proposed on perceptual document analysis to segment text lines [3], or to interpret sketches [4]. In these works the structure of the document is analyzed in terms of salient objects, following perceptual grouping rules. Perceptual organization, i.e. concepts like saliency, closure, repetition, alignment of geometric primitives, etc. seems to be the basis of aesthetics in graphical documents.

Architecture is a visual art where a creative process results in a design that encompasses aesthetics and function. Hence, a composition is influenced by the standards of beauty that can vary depending on the social, cultural and temporal context, but also by the function of the building and its parts. In the conceptualization stage, architects use CAD software or just sketching interfaces for designing new constructions, projecting it in different views, namely the floor plan or the facade. In this stage, a blank paper is considered as a canvas, and the line drawings are like a painting. The language of architecture consists of basic visual tokens as lines, shapes and texture. Lines can be thick or thin, straight or curved, jagged or smooth, dotted or continuous. Shapes define symbols (structure, furniture, utilities). Texture describes materials and object surfaces. The perception of the spatial arrangement and combination of the basic visual elements gives rise to concepts as symmetry, balance, rhythm, proportion, space, etc. A style is characterized by these concepts, influenced by cultural, societal and temporal trends. Roughly speaking, gothic constructions have geometrically ordered and dense ornaments and vertical emphasis; American colonial house plans feature symmetry in doors and windows; Mediterranean style plans include big spaces like patios or courtyards.

This paper is an early attempt to introduce visual aesthetics in graphics recognition, in particular in line drawings of architectural plans in a retrieval scenario. Our hypothesis is that an architect can cast queries in large databases in terms of a sample document and retrieve the architectural drawings from the database having a similar style. The state of the art on floor plan retrieval consists in formulating queries in terms of the functional view of the building. Hence, using ontology models, the user can search for similar designs in terms of the number of rooms, the building symbols, etc. [5]. The approach presented in this paper is appearance-based, so we don't recognize building elements, but correlate perceptual semantic concepts with features extracted from raw images. We focus on two topics, namely spaces and lines. Space refers to the size, the layout, or the shape of rooms. Lines in a broader sense give information of the walls width, texture, or length. Perceptually, spaces and lines give an idea of the building structure (e.g. big squared rooms with thick external walls could be stated as a query in terms of spaces and lines). The structure can be associated to the function of the building (a public theater will have big spaces, a bungalow small ones and thin lines defining the walls).

We propose a very simple image signature model, but descriptive enough to represent semantic topics related to some space and line properties. We propose a runlength histogram descriptor. The intuitive idea is that thick lines representing orthogonal walls generate a high frequency of the codeword of long runs in vertical and horizontal direction, and another high frequency in the codeword of runlenghts corresponding to the width of the walls. On the other hand, big rooms (spaces) are represented by histogram maxima in the runlengths corresponding to their size. The retrieval mode proposed is a query by example framework. Given a query architectural image, we retrieve images that are similar in terms of the space and line topics.

The rest of the paper is organized as follows. In section II we first review the runlength histogram descriptor and afterwards the approach for retrieving architectural drawings is described. In section III the performance evaluation protocol and the experiments are presented. Finally section IV draws the conclusions and outlines the future work.

II. The runlenght histogram representation

A runglenth is the number of linewise consecutive pixels with the same value in a given direction. Given a direction d that defines the scanpath, the *runlength histogram*, denoted as H_d, encodes the length frequencies of the runs in the image extracted linewise at direction d. Thus, $H_d[i]$ counts the number of runs of length i at direction d. Runs can be extracted from foreground or background images. Thus, separate histograms can be considered, or they can be accumulated in a single one depending on the application. To avoid confusions, we will denote as H_d^f, H_d^b or H_d the foreground, background or joint histograms respectively.

Runlength analysis has been used in document analysis with different purposes. Due to the simplicity in the computation, vertical and horizontal runs are the most usually used. In [6], runlength histograms are used to classify between textual and non-textual blocks in administrative documents. Lately, in [7], runlength projections are used to detect page frames in double-page scanned documents. More recently, in [8], the authors suggest the use of runlengths in the detection of potential wall-elements in architectural floor plans. Finally, in [9], runlength histograms are introduced as whole page document description. Here, foreground runlengths at different scales using spatial pyramids are concatenated and used as signature for efficient classification and retrieval tasks in large-scale collections. Strongly based in this idea, we use runlength projections as a perceptual signature model for architectural floor plans.

The pipeline of the process, shown in Figure 1, can be seen as a specific case of existing CBIR strategies. Firstly, all the images are preprocessed for ease of computation. Then, for every floor plan, its signature is calculated by generating the runlength histogram. Given a query example, the method returns a ranked list of the most similar images. Let us further describe the steps of the process.

Preprocess

Floor plans are first binarized using the Otsu method to be able to extract runs. Afterwards, text-graphic separation is performed to filter out text components, because do not appear in all the images and can slightly bias their global perception. In addition to that, images are resized. Thus, the height of the images has been fixed to 1200 pixels meanwhile the width is rescaled dynamically to keep the aspect ratio. With this, we preserve the same image proportions as the ones used to generate the ground-truth, in which the images were shown to the observers in a fixed screen resolution of 1900×1200. The details of the ground-truth generation are explained in Section III-A.

Signature model

The signature model to describe every image is extracted as it is shown in figure 2. A histogram of runs is calculated in the horizontal, vertical and both diagonals for both, foreground and background layers. Histograms are quantized in bins, experimentally determined, distributed in a logarithmic scale as follows:

$$[1],[2],[3-4],[5-8],[9-16],...,[257,-].$$

Then, the signature S of an image j is the concatenation of its directional histograms ordered as follows:

$$S_j = [H^f_{0^\circ} : H^f_{45^\circ} : H^f_{90^\circ} : H^f_{135^\circ} : H^b_{0^\circ} : H^b_{45^\circ} : H^b_{90^\circ} : H^b_{135^\circ}]$$

Each signature contains 2×4×10 = 80 dimensions. Notice that differently than [9], no spatial information is included in the histogram since there is a lack of correlation between the building structure and its location within the floor plan; as common office-documents do (logos and titles at the top, signatures at the bottom, etc). In addition to that, foreground and background histograms are L1 normalized separately. Our intention is to equilibrate the relevancy of the information conveyed by the background and foreground runs, with much higher frequencies in the background ones.

Retrieval

Let P and Q be the signatures of two different floor plan images. Their similarity is calculated by means of the χ^2 distance:

$$\chi^2(P,Q) = \frac{1}{2}\sum_i \frac{(P_i - Q_i)^2}{(P_i + Q_i)}. \quad (1)$$

Given a query instance, the system ranks the rest of the images in the collection according to their affinity to the query, being the first the most similar and the last the most dissimilar.

III. Experiments

A. Formulation of the experimental framework

The starting hypothesis of this work is that similarity between architectural plans can be formulated in terms of perceptually dominant visual cues, without interpreting building elements. The runlength histogram descriptor proposed in this paper is a computational model for the perceptual concepts of space and lines. In a retrieval scenario, given a query floorplan image, the proposed method ranks the database images in terms of the runlength histogram similarity. To validate this output in terms of the visual perception, a ground truth based on subjective human assessment was required. This ground truth was constructed with the participation of human observers that classified images in terms of visual aesthetics. Although it is a subjective assessment, we aimed at

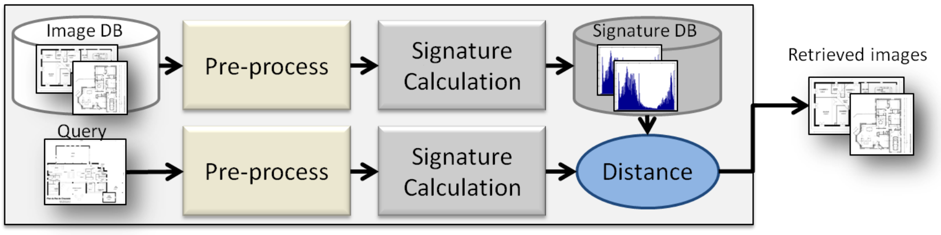

Fig. 1. Pipeline of the method

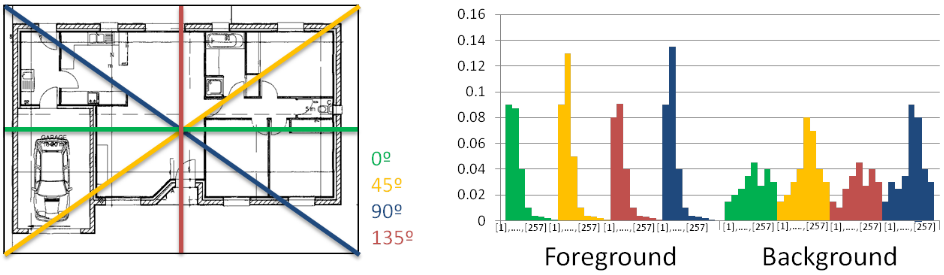

Fig. 2. Signature extraction

statistically corroborate that our computational model validates the hypothesis.

A key issue of the ground truth creation is the design of the procedure to collect the user classification of images. A number of considerations must be made. First, since the goal of the users is to classify images with perceptual features and without interpreting the building elements of the architectural drawings, a pre-attentive experiment was conducted. In cognitive vision, pre-attentive processing is the unconscious accumulation of visual information. To force the users to classify images in a pre-attentive way, they have to be displayed in a short lapse of time (no more than a second). A second consideration in the procedure is the way how images are shown to users. Pairwise learning is usually used in the literature [10]. It consists in decomposing a multi-class classification problem into a set of pairwise problems. Therefore given a query, instead of ranking or classifying the whole set of images it is simpler to compare alternatives in a pairwise way. Finally, a third consideration is the question that is made to the users. They can just be asked to assess if a pair of images are similar or not, or a more focused question that affects the observation of the observer (e.g. ”do you consider that images are similar in terms of space distribution?”, or ”don’t consider the external shape of the plan when assessing the similarity between images”).

Taking into account the above considerations, our ground truth creation was conducted as follows. 20 users participated in the experiment. A database of 39 floor plan images drawn by different architects was used. 6 query images were selected. Each run (different user) showed 234 pairs of images. One of the images was always a query image. Each pair of images was displayed during a second. The user was requested to label each pair of images regarding they were very similar, fair similar or completely dissimilar. At the end, the different labels given by the users to the images where combined averaging the labels so a ranking was generated for each query image.

The human observers selected for generating the ground truth were people aged between 25 and 35 (5 of them were PhD, 14 were PhD students and one had College Studies). None of the volunteers has studied the degree in Architecture and Urban Planning or has knowledge related with floor plans.

We have formulated two different questions to the users. The question **A**, formulated to 10 observers, was: *”Do you consider that these floor plans are from the same architect or architecture studio?”*. The objective of this question was to cluster the floor plans according to the different styles of architectural drawing. The question **B**, also asked to 10 different observers, was: *”Do you consider that these floor plans are similar?”*. Although the question is quite ambiguous and can suggest different criteria, the users were recommended to disregard the shape or the size of the plan. The floor plans of our database have different shapes and sizes/scales but as it has been explained, the goal of the work is to classify them in terms of perceptual cues. This question has similar objective to the first one, but we give more freedom to the users in the task using an unconscious stimuli.

Both formulated questions allowed us to generate three different ground-truth rankings. The *gtA*, taking into consideration the answers from the observers asked with question A. The *gtB*, considering the answers to question B. And *gtAB* considering the answers from all the observers that have participated in the experiment, independently of the question formulated to them.

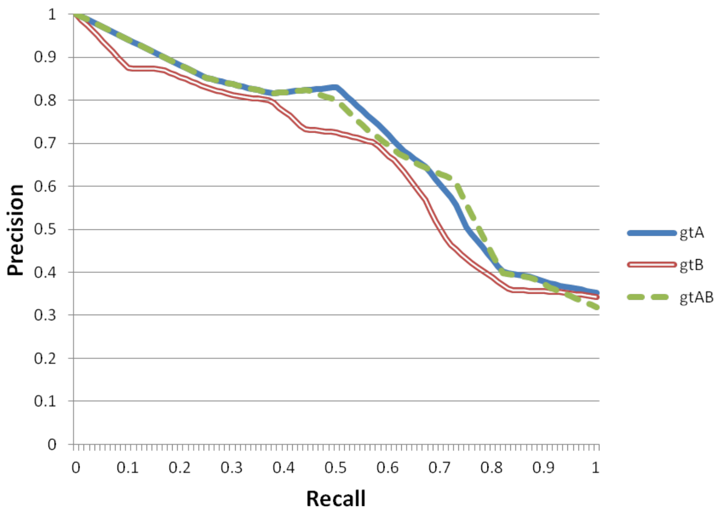

Fig. 3. Precision and recall curves for the three ground-truth models.

TABLE I
PERCEPTUAL RANKING RESULTS FOR THE SIX QUERIES. NOTICE THAT IN THE HYPOTHETICAL CASE OF A RESULTING LIST SORTED INVERSELY TO THE GROUND-TRUTH, THE SCORE WOULD BE 760

	gtA	***gtB***	***gtAB***
image1	115	137	154
image2	64	96	94
image3	139	154	154
image4	33	50	53
image5	110	72	99
image6	123	109	108
mean	**97**	**103**	**110**

B. Performance evaluation protocol

Once the ground truth has been collected after the different user observations as described in the previous section, let us describe how the performance of the proposed approach has been measured. A retrieval process consists in sorting the images of the database according to the similarity to the query. The upper is an image in the ranked list, the more similar is to the query. The ground truth has collected subjective assessments of similarity between the six queries and the database images. These scores given by the users can be turned into ranked lists, so given a query, the performance of the approach is measured in terms of ranked list comparison.

A number of distances between rankings have been proposed in the literature [11]. Let S and S' be two ranked lists of items with the particularity of S' being a permutation of S, i.e. both lists contain exactly the same set of items. Let $\sigma(i)$ denote the rank in S' of the element i in S. The *Spearman's footrule* distance between two rankings measures the total element-wise displacement and is given by:

$$F(S,S') = \sum_i |i - \sigma(i)|;$$

This measure only takes into account the order of the elements to assess the similarity between two lists. The order of the elements is given by their similarity score. Hence, if two elements have the same score value, an arbitrary technique has to decide which element is sorted first. This issue can lead to a random-like ranking when several elements are equally scored in the list and then strongly influence in the ranking distance.

This is the case of our ground-truth ranking. The affine answers obtained from the observers lead to a ranking with many repeated scores for different images. In order to avoid this problem, we have adapted the distance function between two ranking, considering the same distance value for images with the same score. Thus, let S be a ranking obtained by the system, given a query, and S' its corresponding ground-truth, the distance measurement between both rankings is given by:

$$F(S,S') = \sum_i min(|i - \hat{\sigma}(i)|),$$

where $\hat{\sigma}(i)$ denotes the class of equivalence in S' of the element i in S, i.e. a representative among the different images in S' having the same score. Formally:

$$\hat{\sigma}(i) = \arg\min_{j\in S'} \Big(i - \sigma(j) | p_{S'}(j) = p_{S'}(\sigma(i))\Big),$$

where $p_{S'}(j)$ and $p_S(\sigma(i))$ denote the scores of elements j and i of S'.

C. Results and Discussion

In Table I we show the ranking results obtained for six different queries. The numerical values denote the distance between the system retrieval and the three different ground-truth rankings: *gtA*, *gtB* and *gtAB*. The main sorting differences occur in lasts positions of the lists, where most dissimilar images are ranked. Meanwhile, in the first positions, the most similar images are matched for all the queries, a fact that is corroborated by the precision and recall curves extrapolated for all the queries shown in Figure 3. In addition to that, we

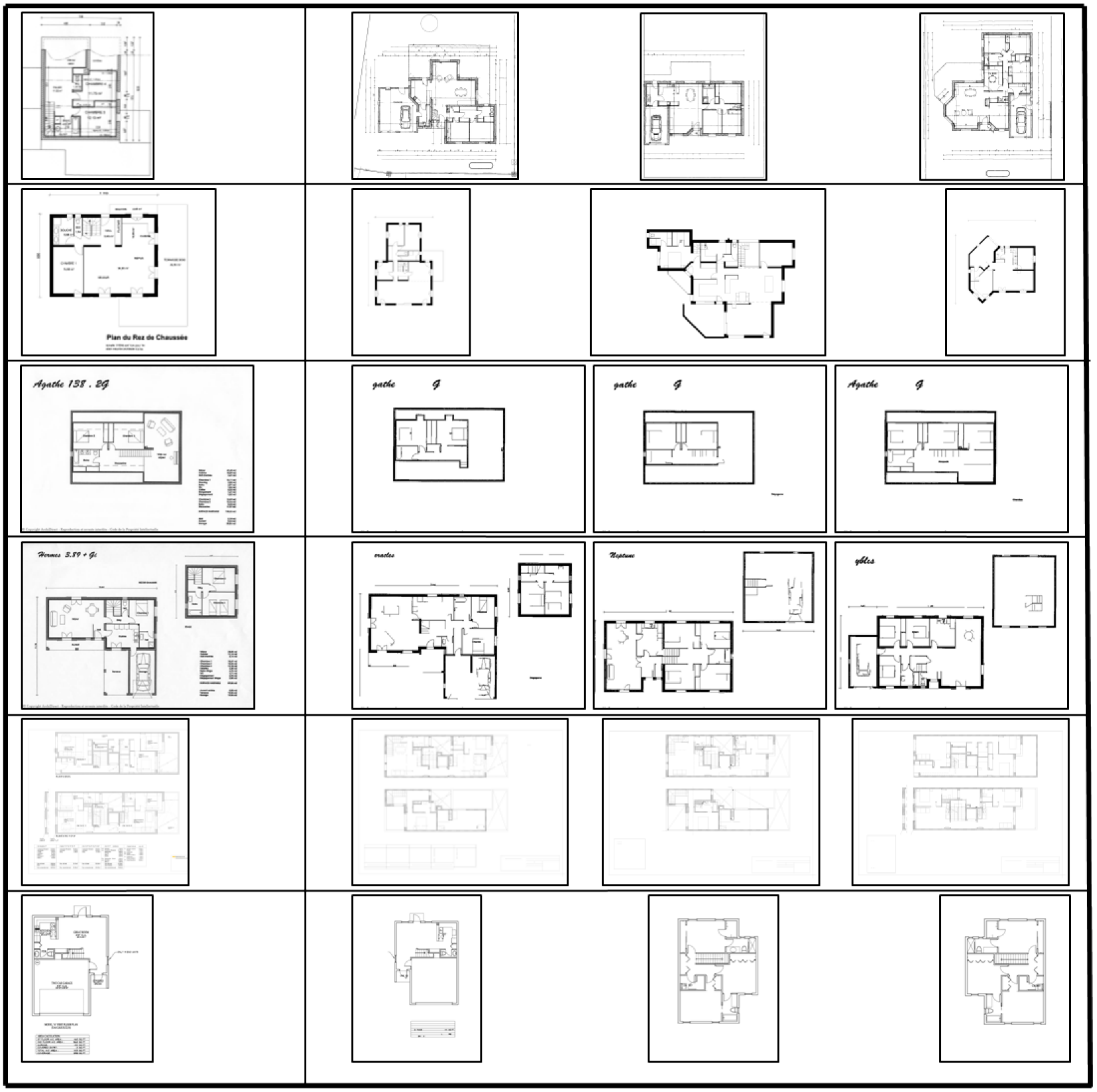

Fig. 4. Perceptual ranking obtained by the system. For each query in the left column, the three most similar images retrieved by the system are shown.

also observe that results are slightly different depending on the groundtruth compared with, being *gtA* the one that leads to the best ranking. In retrieval terms, the mean average precision (mAP) obtained also varies depending on the ground-truth model. Thus, the mAP scores are 0.67 when compared to *gtA*, 0.63 to *gtB*, and 0.66 to gtAB.

In qualitative terms, we show in Figure 4 the six queries and their three respectively most similar images retrieved by the system. As it can be seen in every case, the three retrieved images belong to the same collection of the query –share the same drawing style–, a fact that ratifies the fact just mentioned above regarding the matching accuracy in the first positions of the ranking.

Analysing the experiment performed, we realized about the impact on the system performance of the question formulated to the observers in the ground-truth generation time. It can strongly determine their perception in terms of *similarity*. As it has been shown, different ways to formulate a question aiming for the same answer can imply different interpretations from the observers. Thus, system performance varies when it is compared with *gtA*, *gtB* or *gtAB*. Furthermore, a hight population of observers is needed to generate a trustful ground-truth able to smooth the high impact of biased perceptions from certain individuals.

In global terms and despite its simplicity, the proposed document signature is able to fairly model the human perception of similitude in the floor plans framework.

IV. CONCLUSIONS

In this paper we have proposed a perceptual model to describe aesthetic properties in architectural plans. We have considered that semantic concepts associated to lines and space can make two plans similar to the eyes of a human observer in a pre-attentive process. A simple descriptor based in runlength histograms has been proposed. This signature has the ability to roughly capture the properties of the lines (mainly walls) and the spaces (room aspect) of a floorplan design. In a retrieval framework, this descriptor allows to search for perceptually similar plans into a database. The main contributions of the paper have been: first a pioneer approach of graphical document retrieval based on perceptual cues, without interpreting the building elements, i.e. symbols, of the drawing; and second the design of an experimental setup inspired in the pre-attentive theories of cognitive vision.

Although the work is in a preliminary stage, the obtained results are promising. We have only focused on a few number of perceptual features. A complete system that wants to model the style of an architectural drawing combining function and aesthetics should consider other features representing density, proportion, symmetry, etc. The results presented in section III allow to conclude that run-length histograms capture the intended concepts related to lines and space. When floor plans present orthogonal walls, run-lengths characterizing the width or filling texture of walls present high frequencies. On the other hand, big rooms correspond to long high frequent run-lengths, and small rooms correspond to mid-length ones.

To validate our model, we have collected ground truth based on human observation. A pre-attentive experiment has been conducted, resulting in ranked lists of floor plan images according to the human assessment of their similarity to a given query. We have noticed the subjectivity of this procedure, which corroborates the complexity of human perception in visual aesthetics. We have experimentally observed the relevance of the question made to the users when they are asked to score images.

As a continuation of the work, other features should be considered to describe other semantic concepts. It would allow to have a more complex model of visual aesthetics characterizing styles of architectural drawings.

ACKNOWLEDGEMENT

This work has been partially supported by the Spanish projects TIN2009-14633-C03-03, TIN2011-24631, TIN2012-37475-C02-02, and by the research grants of the Universitat Autonòma de Barcelona (471-01-8/09) and (471-02-1/2010).

REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *Proceedings of the 9th European conference on Computer Vision - Volume Part III*, ser. ECCV'06, 2006, pp. 288–301.

[2] A. K. Moorthy, P. Obrador, and N. Oliver, "Towards computational models of the visual aesthetic appeal of consumer videos," in *Proceedings of the 11th European conference on Computer vision: Part V*, ser. ECCV'10, 2010, pp. 1–14.

[3] A. Lemaitre, J. Camillerapp, and B. Coüasnon, "A perceptive method for handwritten text segmentation," in *Proceedings of Document Recognition and Retrieval XVIII - DRR 2011*, 2011, pp. 1–10.

[4] E. Saund, "Finding perceptually closed paths in sketches and drawings," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 4, pp. 475–491, 2003.

[5] M. Weber, C. Langenhan, T. Roth-Berghofer, M. Liwicki, A. Dengel, and F. Petzold, "a.scatch: Semantic structure for architectural floor plan retrieval," in *Case-Based Reasoning. Research and Development*, ser. Lecture Notes in Computer Science, 2010, vol. 6176, pp. 510–524.

[6] D. Keysers, F. Shafait, and T. M. Breuel, "Document image zone classification - a simple high-performance approach," in *Proceedings of the 2nd Int. Conf. on Computer Vision Theory and Applications*, 2007, pp. 44–51.

[7] N. Stamatopoulos, B. Gatos, and T. Georgiou, "Page frame detection for double page document images," in *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, 2010, pp. 401–408.

[8] L.-P. de las Heras, D. Fernández, 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, p. accepted.

[9] A. Gordo, F. Perronnin, and E. Valveny, "Large-scale document image retrieval and classification with runlength histograms and binary embeddings," *Pattern Recognition*, vol. 46, no. 7, pp. 1898 – 1905, 2013.

[10] J. Fürnkranz and E. Hüllermeier, "Preference learning and ranking by pairwise comparison," in *Preference Learning*. Springer Berlin Heidelberg, 2011, pp. 65–82.

[11] R. Kumar and S. Vassilvitskii, "Generalized distances between rankings," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 571–580.