DIIS - I3A
Universidad de Zaragoza
C/ María de Luna num. 1
E-50018 Zaragoza
Spain

Internal Report: 2007-V09

Door detection in images integrating appearance and shape cues [1]

A.C. Murillo, J. Košecká, J.J. Guerrero C. Sagüés

If you want to cite this report, please use the following reference instead:
Door detection in images integrating appearance and shape cues, A.C. Murillo, J. Košecká, J.J. Guerrero C. Sagüés, *Workshop: From Sensors to Human Spatial Concepts, pages 41-48, IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego- USA, 2007.*

[1]This work was supported by projects DPI2006-07928, IST-1-045062-URUS-STP and NSF Grant No. IIS-0347774.

Door detection in images integrating appearance and shape cues

A. C. Murillo, J. Košecká, J. J. Guerrero and C. Sagüés

Abstract

Important component of human-robot interaction is the capability to associate semantic concepts to encountered locations and objects. This functionality is essential for visually guided navigation as well as place and object recognition. In this paper we focus on the problem of door detection using visual information only. Doors are frequently encountered in structured man-made environments and function as transitions between different places. We adopt a probabilistic approach to the problem using a model based Bayes inference to detect the door. Different from previous approaches the proposed model captures both the shape and appearance of the door. This is learned from a few training examples, exploiting additional assumptions about structure of indoors environments. After the learning stage, we describe a hypothesis generation process and several approaches to evaluate the probability of each generated hypothesis. The new proposal is tested on numerous examples of indoor environments, showing a good performance as long as enough features are encountered.

I. Introduction

In this paper we present a new technique for detecting doors in perspective images of indoors environment using only visual information. Detection of doors is of great importance for various navigation and manipulation tasks. The doors are often places which separate different locations, can be used as landmarks for navigation and/or relative positioning or as hints to define exploration and SLAM strategies [1]. They also need to be recognized for door opening and navigation to neighbouring rooms [2], [3].

The problem of door detection has been studied numerous times in the past. The existing approaches differ in the type of sensors they use and the variability of the environment/images they consider. For example, doors are detected using both visual information and range data (sonar) in [4] and [5]. In [4] authors exploit the fact that vision is good for providing long range information (beyond the range of ultrasound sensor) and detect and group the vertical lines based on the expected dimensions of the door and form initial door hypotheses. In [5] the authors tackle more general problem of obtaining a model of the environment defined by instantiations of several objects of predefined class (e.g. doors, walls) given range data and color images from an omni-directional camera. The doors are then detected as particular instantiations of the door model, given all the sensory data. The door hypotheses are obtained by fitting linear segments to laser range data and associated color values from the omnidirectional camera. In [6] authors focus on handling the variations in door appearance due to camera pose, by characterizing properties of the individual segments using linguistic variables of size, direction and height and combine the evidence using fuzzy logic. Additional work using visual information only was reported in [7], where only geometric information about configurations of line segments is used. In most instances, only the doors which were clearly visible and close to the observer were selected as correct hypotheses.

Additional motivation for revisiting the door detection problem is to explore the suitability of general object detection/recognition techniques to door detection/recognition problem. A more general goal is to use the obtained insights for tackling recognition and detection of additional objects in the environment, which do not have a very discriminative appearance and are to a large extend defined by their shape, but whose shape projection varies dramatically as a function of viewpoint, e.g. tables or shelves. The object recognition techniques, which are being explored extensively in computer vision, commonly adopt so called part based models of the objects, which consider representations of objects in terms of parts [8] and spatial relationships between them. Learning the object parts for different object classes is often the first stage of existing approaches. The classification methods then vary depending whether full generative model is sought or discriminative techniques are used, or combination of both. In the simplest of the generative model settings, the recognition stage proceeds with the computation $p(Object|X, A)$, where X, A are the positions and appearance of the object parts and $p(X, A|\theta)$ is learned from training examples, where θ are the parameters chosen to describe the object. With the discriminative approaches, multi-class classifiers are trained to distinguish between the low-level features/parts characteristic of a particular class [9] and typically proceed in a supervised or weakly supervised setting. In the robotic domain the discriminative approach has been applied for place and door recognition using Adaboost learning procedure, with geometric features computed from laser and Haar-like features computed from images as input features [10].

In this work, we explore the Bayesian approach, where we learn the probability distribution $P(Door|\theta)$ in a parametric form in a supervised setting. Alternatively to the part based representation assumed in [8], we pursue a model based approach, where the geometry of the door is specified by small number of parameters and appearance is learned from a few training examples. This type of representations resemble models used in interpretations of architectural styles and man-made environments, where the analysed scenes are typically well characterized by a small number of geometric/architectural primitives [11]. Instead of proposing the generative model of the whole image, we use the constraints of man-made environments to reduce the search space and to generate multiple hypotheses and use the learned probability distribution to evaluate their likelihood.

A. C. Murillo, J.J. Guerrero and C. Sagüés are with the I3A/DIIS at University of Zaragoza, 50010 Spain. acm@unizar.es
J. Košecká is with the Department of Computer Science George Mason University, Fairfax, VA, 22030, USA. kosecka@cs.gmu.edu
This work was supported by projects DPI2003-07986, DPI2006-07928, IST-1-045062-URUS-STP and NSF Grant No. IIS-0347774.

Outline

Section II describes the probabilistic model we adopt. The hypothesis generation process is explained in Section III, followed in Section IV by the approaches for learning the appearance likelihood of the model. Section V explains how the likelihood evaluation of each hypothesis is performed. Finally sections VI and VII present initial door detection experiments and some conclusions of the work.

II. Problem formulation

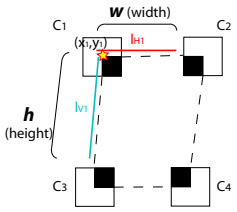

Fig. 1. Model of a door and components of one of its corner-features (C_1): corner location (x_1, y_1) and lines that give rise to the cross point (l_H, l_V).

We will assume that the door model, see Fig. 1, is well described by a small set of parameters θ. Ideally, if we were to pursue a fully Bayesian approach, we would first learn or have at our disposal prior distributions of these parameters. We start with a restricted simple setting where we seek to compute $p(Object|X, A)$, with X, A being the positions and appearance of low-level features detected in the image:

$$p(Object|X, A) \approx p(X, A|Object)P(Object).$$

Assuming that all objects are equally likely and that our object of interest can be well described by a small set of parameters $\theta = (\theta_S, \theta_A)$, shape and appearance parameters respectively, this posterior probability can be decomposed:

$$\begin{aligned} P(\theta|X, A) &\propto P(X, A|\theta)P(\theta) \\ &= P(X, A|\theta_S, \theta_A)P(\theta_S, \theta_A) \\ &= P(X, A|\theta_S, \theta_A)P(\theta_A)P(\theta_S). \end{aligned} \tag{1}$$

We consider the parameters θ_S and θ_A to be independent, e.g. appearance (color/texture) of a primitive is independent of its shape and vice versa. The interpretation of the final terms in (1) is as follows:

- $P(\theta_S)$ represents the prior knowledge about the geometric shape parameters of the door, for instance the ratio between width and height of the door or the position of the c_3 corner, which should be touching the floor.
- $P(\theta_A)$ is the prior information on the appearance of the object, in our case doors. This information is typically learned from examples. In this work we will exploit only color information, but more elaborate appearance models based on texture can be incorporated.
- $P(X, A|\theta_S, \theta_A)$ is the likelihood term of individual measurements, given a particular instantiation of the model parameters $\theta = (\theta_S, \theta_A)$. In the presented work, we do not use priors $P(\theta_A)$ and $P(\theta_S)$ and consider only maximum likelihood values of the parameters θ_S and θ_A, which for geometry are given by a known model and for appearance are learned in a supervised setting. The likelihood term can be further factored as

$$P(X, A|\theta_S, \theta_A) = P(A|X, \theta_S, \theta_A)P(X|\theta_A, \theta_S).$$

The shape likelihood term used in this work is explained in Section V-B and the possible choices for the appearance likelihoods are described and evaluated in Sections IV and V-A.

III. Hypotheses generation

The selection of individual hypotheses consists of a low level feature extraction process followed by the initial hypotheses instantiation which we describe next.

A. Feature extraction

First, line segments are extracted from the image with our implementation of the reference [12], and the vanishing points are estimated with the approach described in [13]. Using vanishing point information the line segments are grouped in two sets: lines which are aligned with the vertical vanishing direction and lines which are aligned with either horizontal direction or the z optical axis. All possible intersections between vertical and the remaining sets of lines are computed. The intersection points which have low corner response (measured by Harris corner quality function) are rejected. Figure 2 shows an example of the extracted lines grouped using the vanishing information (in red vertical ones, in blue non-vertical ones). In the same figure, all the intersection points that were close to line segments are shown with a cross (+), and those that remained after the high cornerness response filtering are re-marked with squares around (□). Finally the detected intersections are classified into 4 types (c_1, c_2, c_3 and c_4), according to the kind of corner that they produce (see Fig. 3).

B. Instantiation of initial hypotheses

The corner features detected in the previous stage are grouped into sets of compatible ones, which are used to define initial hypotheses. In the first stage pairs of compatible corners ($\{c_1, c_2\}$, $\{c_1, c_3\}$, $\{c_2, c_4\}$ and $\{c_3, c_4\}$) are found. To consider a pair of corners to be compatible we take into account its alignment, according to the directions of the lines (l_V,l_H) that generated those corner features. For example, a corner of type c_1 is considered compatible with all corners of type c_2 which are on the right of the c_1 corner and whose respective line segments l_H are aligned up to a small threshold. This search for two corner hypotheses is followed by the intersection between these sets of 2 corners, obtaining sets of 3 compatible corners: $\{c_1, c_2, c_3\}$, $\{c_1, c_3, c_4\}$, $\{c_1, c_2, c_4\}$, $\{c_2, c_3, c_4\}$. Similarly, we look for intersections between the 3-corner hypotheses to obtain hypotheses supported by 4-corners $\{c_1, c_2, c_3, c_4\}$.

After this stage, we have four types of hypotheses: supported by 4, 3 or 2 corner features or composed by those singleton corners that did not have compatible corner features. Example hypotheses generated for an image are shown in Fig. 4. All the corner features extracted are marked by a square (□), the corners contributing to each hypothesis are marked by $*$, and the dotted lines show the area delimitated by the hypothesis. Each subplot shows the hypotheses contributed by 1, 2, 3 or 4 corners respectively for the same test image. Only for the 4 corner hypotheses all supporting corners correspond to real corner features (□). In the remaining cases the missing corners are generated by completing the rectangle with the information from the available corner features, using their supporting lines as shown in Fig. 5.

IV. Learning the appearance model

The appearance model of the door is learned in a supervised setting from a set of examples. In this work, the appearance model was learnt in two steps: first a set of training images was segmented based on color and individual segments were labeled as door or non-door segments. Then, we investigated various means of representing the appearance (in our case color) probabilistically.

Fig. 2. Line segments grouped in vanishing directions (vertical in red, non-vertical in blue), corner points detected (green +) and corner points with high corner response (black □).

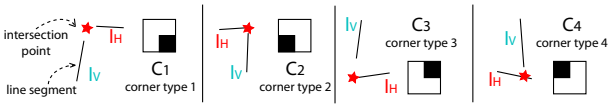

Fig. 3. Examples of line intersections and the four types of corner features that can be generated.

Fig. 4. Initial hypotheses generated for a test image. All extracted corner features are shown with a black square. Top-left: 1-corner hypotheses; Top-right: 2-corner hypotheses ; Bottom-left: 3-corner hypotheses; Bottom-right: 4-corner hypotheses

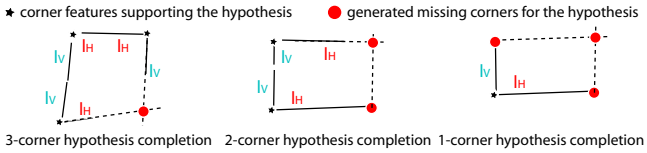

Fig. 5. Examples of the completion of initial hypothesis supported by less than 4 corner features.

A. *Reference images segmentation and labeling*

The set of training images was segmented using the color based segmentation algorithm proposed in [14] at different levels of accuracy. All obtained door segments were labelled. Fig. 6 shows two examples of reference images segmented at the finest level tried ($\sigma = 0.5;\ k = 500;\ min\_size = 20$) where k is the maximum number of segments, while minimal segment size is 20 pixels. The value of σ is used in the smoothing stage preceding the algorithm. Using this level of segmentation generated segments of desirable size, without including too many pixels outside of door regions.

Fig. 6. Two training images and their segmentations. Left: the original images. Right: segmentation obtained (the dotted pattern marks the segments labelled as doors).

B. *Representations of the appearance information*

Once the door-segments are detected and labelled, the appearance properties of the modeled object are learned. Fig. 7 shows the door and non-doors pixels from the reference images plotted in the RGB and Lab (CIE 1976 (L* a* b*)) color spaces. These distributions were studied for different levels of segmentation. Not surprisingly for coarser segmentation levels, the

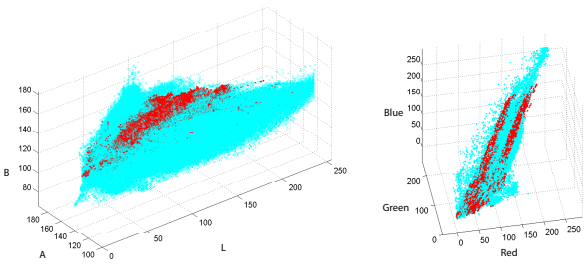

Fig. 7. Reference door (red) and non-door (blue) pixels plotted in Lab (left) and RGB (right) color spaces.

clusters belonging to doors were less compact. The color space Lab represents the pixels from the labelled segments in a more compact way.

In order to be able to compute the appearance likelihood $P(A|X,\theta_A)$ of pixels in a particular region to be door pixels, we need to adopt a particular form of probability density function or alternative means of computing the probabilities directly. Here we examine three such choices and discuss their advantages and disadvantages:

1 - The first approach examined was motivated by earlier work on using color for face detection [15]. Here instead of assuming a parametric model of pdf, authors used directly the data (pixels counts) acquired in the training stage. In this setting, probability that a given rgb pixel value x belongs to a door is obtained as

$$P(x|\theta_A) = \frac{r(rgb)}{t(rgb)}, \tag{2}$$

where $r(rgb)$ is the count of pixels with color rgb that were labeled as a door segments, and $t(rgb)$ is the total count of pixels with color rgb in the reference images.

Fig. 8 shows several examples of the probability assigned to each pixel following this approach.

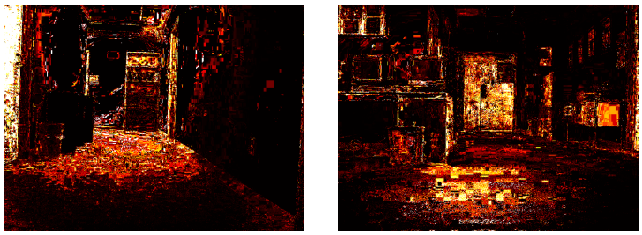

Fig. 8. Examples of the probability assigned to each pixel based on the reference pixels counts.

2 - The color distribution of the pixel is modeled as a mixture of Gaussians. The probability of a particular rgb pixel value x then takes the following form

$$P(x|\theta_A) = \sum_{j=1}^{k} \alpha_j \frac{1}{\sqrt{(2\pi)^d|\Sigma_j|}} e^{\frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)} \tag{3}$$

where $\alpha_j > 0$ and $\sum_{j=1}^{k} \alpha_j = 1$ and $\theta_A = \{\mu_i, \Sigma_i\}_1^k$ are the means and covariances of learned clusters. Instead of using the entire probability distribution to determine the probability of a pixel having a door appearance, we find the Gaussian with the highest probability and assign that probability to the pixel

$$P(x|\theta_A) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_j|}} e^{\frac{1}{2}(x-\mu_j)^T\Sigma^{-1}(x-\mu_j)}. \tag{4}$$

Similarly to RGB color representation, the same model can be adopted with any other three band color representations, such as HSV or Lab. Figure 9 show the clustering results for RGB and Lab color spaces. They correspond to the finer level of segmentation and also finer clustering (k=10), although tests were done also with coarser levels of segmentation and clustering. Fig. 10 shows several examples of the appearance likelihoods assigned to individual pixels following this approach.

3 - The mixture of Gaussians model has been also applied to color histograms of door regions. In this case, for each labeled segment in the training set, a histogram of the color distribution is built. Two ways of building these histograms have been studied. First, only *marginal histograms* were considered and each color band was quantized in 50 bins, therefore each region is represented by a 150 bins histogram. With this representation we are assuming that the three color bands are independent, yielding a low dimensional representation of the histogram. This assumption has been successfully applied before [16] and has been shown to be useful in cases where there are few training examples (as it occurs in our case).

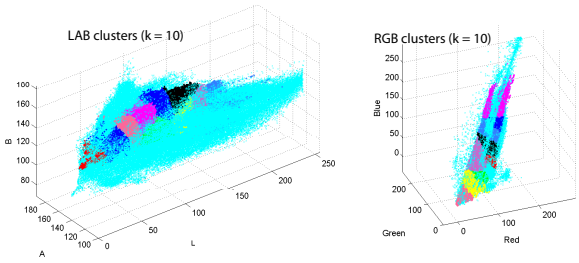

Fig. 9. K-means clustering of color from labelled door pixels in Lab (left) and RGB (right) color spaces (the big set of blue + points corresponds to non-door pixels).

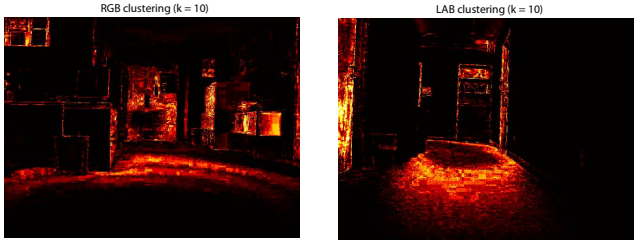

Fig. 10. Examples of the probability assigned to each pixel based on the mixture of Gaussians.

In the second approach, *full histograms* were considered. The color space is quantized from 24 bits (considering 256 possible values for each of the three color bands) to 12 bits (2^4 possible values for each band). Then each of the three color bands should be fitted to a range between 0 and 15, giving a range of 4096 possible colors. Each color rgb now is represented as $256 * \frac{r}{15} + 16 * \frac{g}{15} + b + 1$ and 4096 bin histogram can be computed for each region using this representation.

The generated histograms, with either of these two approaches are clustered with k-means and each cluster is represented by a centroid, mass and covariance. The probability of a certain region determined by features X having an appearance of a door depends on the distance, d_h, to the closest cluster:

$$P(A|X, \theta_A) = e^{\frac{-d_h}{\sigma_h}}. \tag{5}$$

The distance d_h between two normalized histograms $h1$ and $h2$ is based on the Bhattacharyya distance, $d_h(h1, h2) = 1 - \sum_{i=1}^{n} \sqrt{h1(i)h2(i)}$, where n is the number of histogram bins. σ_h is the deviation among this distance computed from the query to door and non-door reference histograms.

V. Likelihood computation

The previous section describes several door appearance models, which can be used to compute the probability of a pixel or a region of pixels having a door appearance. In case of analytical distribution models we have shown how to determine the parameters of the model θ_A in the supervised learning stage and consequently determine probability of appearance $P(A|X, \theta_A)$ given the appearance parameters θ_A and shape of the region X. In this section we will demonstrate the computation of the complete likelihood $P(X, A|\theta_S, \theta_A)$ for each of the initial hypotheses. Lets suppose the measurements D for all the generated hypotheses are characterizing some shape attributes X and some appearance attributes A. In order to compute the probability of $P(D|\theta)$ we have

$$\begin{aligned} P(X, A|\theta) &= P(X, A|\theta_A, \theta_S) \\ &= P(A|X, \theta_A, \theta_S)P(X|\theta_A, \theta_S) \\ &= P(A|X, \theta_A)P(X|\theta_S). \end{aligned} \tag{6}$$

A. Evaluation of the appearance likelihood

First, the appearance likelihood of each hypothesis given its shape $P(A|X, \theta_A)$ is evaluated. Different strategies are examined, depending on the representation of the appearance/color distribution discussed in the previous section:

1 - In the case of the first representation we simply integrate probabilities of all pixels in the region delimited by the hypothesis. Each pixel color is given a probability, based on the count of each color in the reference segments, as explained

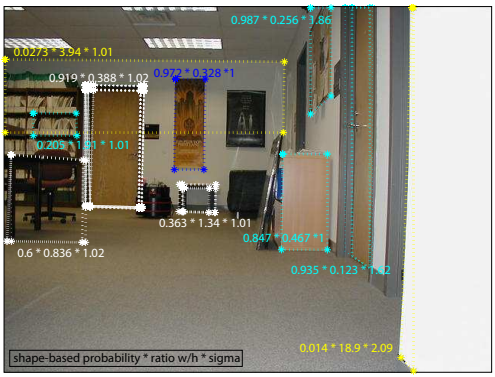

Fig. 11. Some examples of hypotheses with its $P(X|\theta_r)$ (see eq. 9), and corresponding ratio $\frac{w}{h}$ and σ.

in eq. (2) from Section IV. The appearance likelihood estimation takes into account all pixels in the hypothesis region with individual probability higher than a threshold τ:

$$P(A|X,\theta_A) = \frac{\sum_X P(x|\theta_A) > \tau}{c(X)}, \tag{7}$$

where $c(X)$ is the number of pixels in the rectangular region X delimiting the hypothesis.

2 - When the pdf is modelled using Gaussian Mixture Model in different color spaces, individual pixel likelihoods, $P(x|\theta_A)$, are computed based on the distance to the closest cluster (using RGB and Lab spaces), as described in eq. (4). The entire likelihood of the hypothesis is then computed with the same eq. (7) above.

3 - In the case of histogram representation, the pixels do not get an individual color based likelihood, but histograms computed from the region determined by each hypothesis are used. The probability of the hypothesis being correct is equal to the probability of the histogram that represents it. A histogram has a likelihood obtained from its smallest distance to a reference door histogram cluster (d_h), as explained before in eq. (5) of Sec. IV.

B. Evaluation of the shape likelihood

The shape model of our object of interest can be characterized by parameters $\theta_S = [n_c, r]$, where n_c is the number of corner features that support the model and $r = \frac{w}{h}$, being width w and height h the dimensions of the object (door) perimeter (see Fig. 1). The likelihood of a hypothesis given its shape parameters is a combination of two terms

$$P(X|\theta_S) = P(X|\theta_{n_c})P(X|\theta_r),$$

where X are the shape related measurements associated with the hypothesis, namely the supporting corners and the shape of the associated region. The first term, $P(X|\theta_{n_c})$ assigns higher likelihood to hypotheses which were supported by larger number of corner features. It consists of the discrete pdf described in next eq. (8):

$$P(X|\theta_{n_c}) = \begin{cases} 0.7 \text{ when } n_c = 1 \\ 0.8 \text{ when } n_c = 2 \\ 0.9 \text{ when } n_c = 3 \\ 1 \text{ when } n_c = 4. \end{cases} \tag{8}$$

The second term $P(X|\theta_r)$ takes into account the ratio between the height and width of a door. We consider a typical ratio $\frac{w}{h}$ for a common door in a frontal view as 0.3. Then we evaluate the ratio between the height and width of the hypothesis taking into account how far is that hypothesis from the frontal view of a door by checking the perpendicularity of the lines composing the corner features. Then the shape likelihood terms becomes

$$P(X|\theta_r) = e^{\frac{-(\frac{w}{h}-0.3)}{\sigma^2}}, \tag{9}$$

with $\sigma^2 = \frac{\pi/2}{\alpha}$, where α is the angle, range $[0\ , \pi/2]$, between non-vertical line and vertical line that determined a certain corner. It takes into account the variability on the confidence of the ratio as a function of viewpoint. In Fig. 11, there are several examples of hypotheses supported with different number of corners. It is shown for each hypothesis the probability obtained with eq. (9) followed by the ratio $\frac{w}{h}$ and the sigma obtained as explained above. It can be seen there, how this probability evaluation gracefully handles perspective distortion, except when the hypothesis has really bad ratio (see yellow hypothesis on the right of the figure). When the hypothesis is close to a fronto-parallel view, the ratio between width and height correctly penalizes ”non-door” rectangular hypotheses.

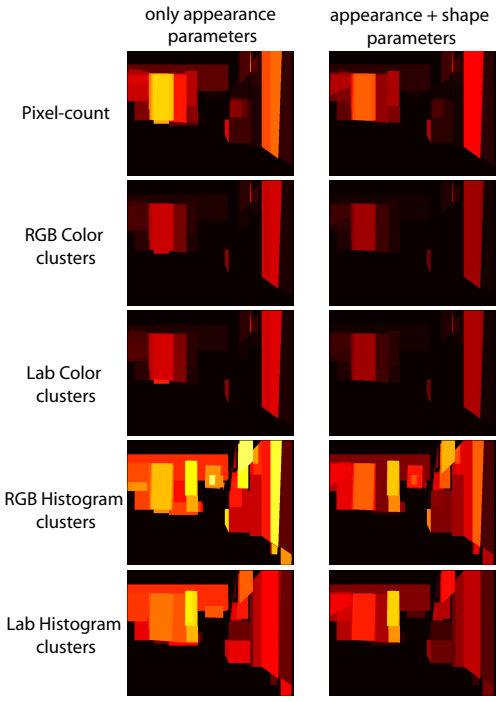

Fig. 12. Left: probability assignment evaluating only the appearance parameters of each hypothesis (with different approaches). Right: same examples evaluating shape parameters as well. (The original test image is the same as in Fig. 11).

The probability of a hypothesis given a certain set of shape parameters can be integrated with the appearance-based probability obtained previously (section V-A). Fig. 12 shows some examples where it is possible to observe the improvement in the probability assignment to the hypothesis when the shape-based prior probability, regarding the ratio between the width and height of the hypothesis, is included. The left column of the figure shows some posterior probability assignments without it, and the right column, the same cases but with the shape information included.

VI. Experimental Results

This section shows the most representative results from an extensive set of experiments, pointing the advantages and disadvantages of the best performing approaches studied.

The experiments were performed with conventional images from two different sets, corresponding to a robot tour around two indoor office like environments. Only four images from each set were used for learning the appearance model, (following the process for segmentation, labelling and learning explained in Section IV). The four images from one set contained 3 frontal views of doors and 3 oblique views of doors. The four images from the other set contained 9 oblique views. All doors are wooden ones, except two of them that are elevator doors. Around half of the remaining images (37 images, with 76 doors) distributed along the two robot tours were used to test the performance of the door recognition approaches.

Fig. 13 shows an example of the likelihood evaluation with different approaches for a test image. First row images show the features extracted, the initial hypotheses and the likelihood evaluation obtained if only the shape likelihood term (θ_S) is taken into account. The second row shows the likelihood evaluation if only the appearance likelihood term (θ_A) is taken into account, and the third one the results of using the overall likelihood in the evaluation. Notice the improvement of integrating both shape and appearance cues for the likelihood evaluation with any of the appearance representations studied.

In order to evaluate the door recognition performance, different criteria could be followed to decide which hypothesis are accepted as doors after the likelihood evaluation. As most hypotheses get a not-null probability of being a door, the most suitable approach is to establish a threshold to accept or reject hypothesis. Fixed and variable thresholds have been studied. The variable ones, depending on the likelihood evaluation of the current image, have shown better performance (e.g. taking as acceptance threshold the median value of the likelihoods or a 75% of the maximum likelihood in the current test image). Fig. 14 shows a typical test image with the hypothesis accepted using two different criteria (thresholds) and their estimated likelihoods.

The summary of the different approaches performance is shown in Fig. 15. Each three-bar plot represents the percentage of doors recognized ($\%\ recognized$), the percentage of door reference pixels included in the accepted hypothesis ($\%\ pixels\ ok$) and the percentage of the non-door reference pixels included in the accepted hypothesis ($\%\ pixels\ false\ positives$). The best results in all cases were obtained using as acceptance threshold the median. The results with the color clustering approach

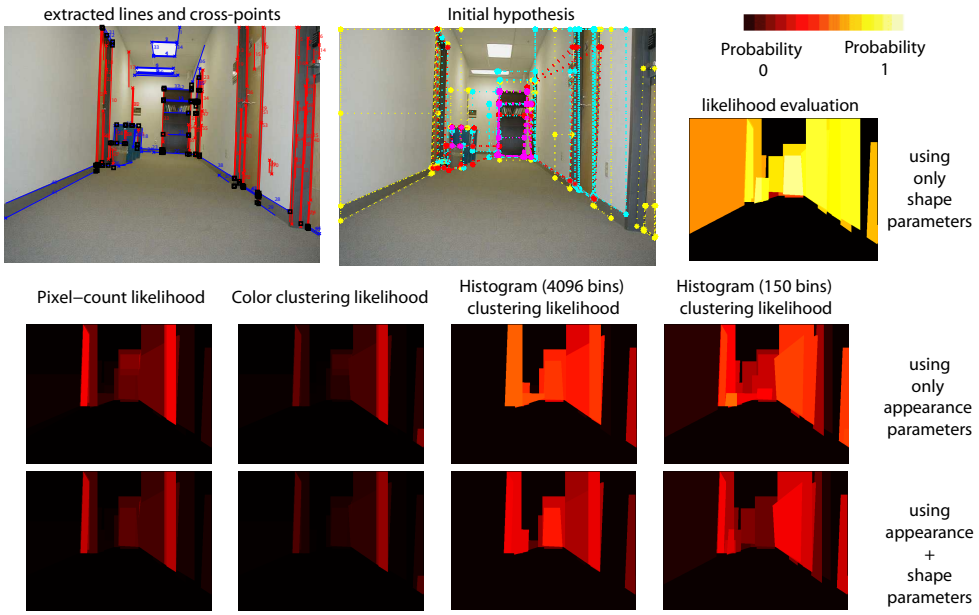

Fig. 13. Likelihood evaluation with different approaches for a test image in the Lab color space.

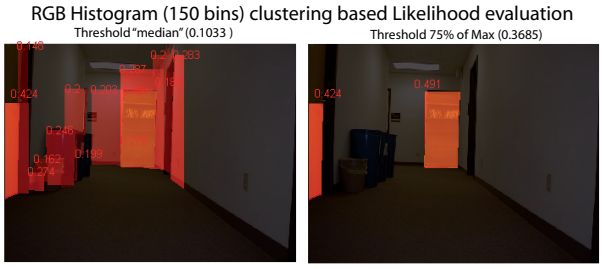

Fig. 14. Hypothesis accepted for a test image under different approaches/criteria in the RGB color space.

were quite similar using the Lab color space and the RGB space, and similarly occurred for the *marginal histograms* (150 bins) clustering. However, using the *full histogram* approach (4096 bins histograms) the Lab color space gave better results.

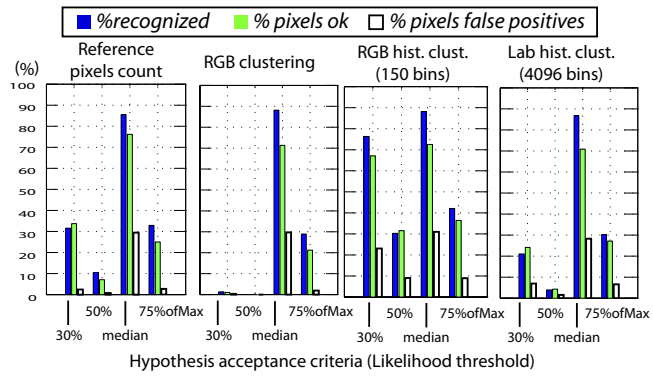

Fig. 15. Door recognition performance for the main approaches studied.

In all cases, the best results were obtained using the median as acceptance threshold. Fig. 16 shows the ROC curve for the same experiments than in Fig. 15, with the relationship between true and false positives using the range of percentiles from 0 to 100 as acceptance thresholds. We can observe again similar behaviour for all cases.

Table I presents more detailed information of the approaches with better performance. Each column corresponds to a different approach: RGB clustering, RGB Histograms (of 150 bins) clustering and Lab Histograms (of 4096 bins) clustering. $\#doors$ is the total amount of doors in the test images, row ok shows the percentage of doors recognized and row $pix.ok$ the percentage of door pixels included in the accepted hypothesis. $\%\ ok$ is always higher or equal than $\%\ pix.ok$, as accepted door hypothesis usually cover a smaller region in the image than the manually selected region for the evaluation. Rows $front.$, $whole$ and

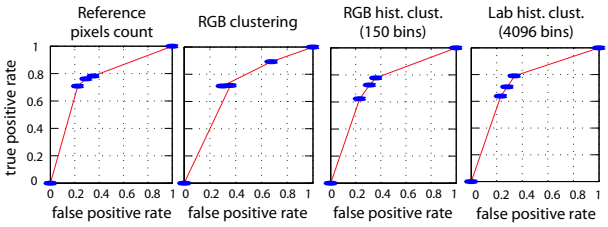

Fig. 16. ROC curve for the same experiment than in Fig.15. It shows the relationship between true and false positives when the acceptance criteria varies using different percentiles as thresholds. The five measurements in each plot (marked by a blue ellipse) correspond to the percentiles 0, 25, 50, 75 and 100 used as acceptance threshold.

$closed$ show the recognition rate if only the corresponding kind of door view (frontal view, whole door view or closed door) is taken into account. Finally row $pix.fp$ contains the percentage of non-door pixels included in the accepted hypothesis (false positives), and $noDoor.fp$ is the same measurement but in the special case of images with no doors, to evaluate how the process performs in that situation. All approaches give a relatively high rate of false positives, mainly due to test images with no doors, where the variable thresholds do not seem very suitable. To try to deal with these cases, the same experiments were run with a minimum fixed global threshold of 5% in addition to the variable one. This decreased the number of false positives, especially with the RGB clustering approach (from 30% $pixels.fp$ to 6% and from 37% $noDoor.fp$ to 0.006%), but also the number of true positives (from 88% ok to 46%). The histogram clustering based approaches kept very similar results with this minimum fixed threshold.

TABLE I

DOOR RECOGNITION EVALUATION (*median* AS ACCEPTANCE CRITERIA)

$\#doors : 76$	ok	$front.$	$whole$	$closed$	$pix.ok$	$pix.fp$	$noDoor.fp$
RGB clust.	88%	83%	84%	90%	71%	30%	37%
RGB Hist.(150)	88%	92%	78%	90%	73%	31%	34%
Lab Hist.(4096)	87%	83%	76 %	88%	71%	28%	35%

Note that most of the non-recognized doors were too far, therefore too small to extract features there, or they were highly occluded and no corner-features were extracted in the door region. Fig. 17 shows some of these non-recognized doors and Table II the results for the same experiments detailed in previous Table I if we do not take into account the doors smaller than a threshold (in this case with image area smaller than 2000 pixels).

Then, the main reason of failure in the recognition seems to be the lack of features for very small (distant) and narrow doors, so this could be improved by adding alternative hypothesis instantiation possibilities where no cross points from lines would be necessary or considering alternative models of the shapes. The high percentage of false positives obtained with some approaches may be improved by adopting more complex appearance representation.

TABLE II

DOOR RECOGNITION OF DOORS OVER 2000 PIXELS IN THE IMAGE

$\#doors : 51$	ok	$front.$	$whole$	$closed$	$pix.ok$	$pix.fp.$	$noDoor.fp$
RGB clust.	92%	83%	89%	93%	71%	31%	27%
RGB Hist.(150)	90%	92%	82%	93%	70%	31%	33%
Lab Hist.(4096)	88%	83%	79%	90%	68%	28%	31%

Fig. 17. Typical failure examples: the ellipses point some typical doors that are not correctly identified with any approach.

VII. Conclusions

In this paper we have presented a new technique for detecting doors using only visual information. The probability distribution $P(Door|\theta)$ is learnt in a parametric form from a few reference images in a supervised setting. A model based approach is taken, where the door model is described by a small set of parameters θ characterizing the shape and the appearance of the object. The geometry of the door is specified by small number of parameters and the appearance is learned from the reference data. We use constraints of man-made environments to generate multiple hypotheses of the model and use the learned probability distribution to evaluate their likelihood. The approach has been extensively tested and evaluated, with good recognition rates as long as the process is able to detect some of the door features.

As future work, we plan to investigate alternative appearance models and incorporate some priors on the shape parameters, e.g. a door should be touching the floor. With more complex models, this approach could be easily extended to a more general setting and allow to explore possibilities to disambiguate between other objects with not very discriminative appearance and large shape distortions induced by change of viewpoint, such as tables or shelves.

References

[1] N. Tomatis, I. Nourbakhsh, and R. Siegwart. Hybrid simultaneous localization and map building: a natural integration of topological and metric. *Robotics and Autonomous Systems*, 44:3–14, 2003.

[2] R. Brooks, L. Aryananda, A. Edsinger, P. Fitzpatrick, Ch. Kemp, U.-M. O'Reilly, E. Torres-Jara, P. Varshavskaya, and J. Webber. Sensing and manipulating built for human environments. In *Int. Journal of Humanoid Robotics*, volume 1, pages 1–28, 2004.

[3] S. Vasuvedan, S. Gachter, V. Nguyen, and R. Siegwart. Cognitive maps for mobile robots - an object based approach. *Robotics and Autonomous Systems*, 55(5), 2007.

[4] S.A. Stoeter, F. Le Mauff, and N. P. Papanikopoulos. Real-time door detection in cluttered environments. In *2000 Int. Symposium on Intelligent Control*, pages 187–192, 2000.

[5] D. Anguelov, D. Koller, E. Parker, and S. Thrun. Detecting and modelling doors with mobile robots. In *IEEE Int. Conf. on Robotics and Automation*, pages 3777–3784, 2004.

[6] R. Muñoz-Salinas, E. Aguirre, M. Garcia-Silvente, and A. Gonzales. Door detection using computer vision and fuzzy logic. In *WSEAS Transactions on Systems*, pages 10(3):3047–3052, 2004.

[7] W. Shi and J. Samarabandu. Investigating the performance of corridor and door detection algorithms in different environments. In *Int. Conf. on Information and Automation*, pages 206–211, 2006.

[8] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 264–271, 2003.

[9] A. Torralba, K. Murphy, and W. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2004.

[10] C. Stachnis, O. Martinez-Mozos, A. Rottman, and W. Burgard. Semantic labeling of places. In *Int. Symposium on Robotics Research*, 2005.

[11] A. Dick, P.Torr, S. Ruffle, and R. Cipolla. Combining single view recognition and multiple view stereo for architectural scenes. In *IEEE Int. Conf. on Computer Vision*, volume I, pages 268–274, 2001.

[12] J.B. Burns, A.R. Hanson, and E.M. Riseman. Extracting straight lines. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(4):425–455, 1986.

[13] J. Kosecka and W. Zhang. Video compass. In *Proc. of European Conference on Computer Vision*, pages 657 – 673, 2002.

[14] Pedro F. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, 2004.

[15] M. J. Jones and J. M. Rehg. Statistical color models with application to skin detection. *Int. J. Comput. Vision*, 46(1):81–96, 2002.

[16] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *Comput. Vis. Image Underst.*, 84(1):25–43, 2001.