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

Internal Report: 2007-V04

Visual Door Detection 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:
Visual Door Detection Integrating Appearance and Shape Cues, A.C. Murillo, J. Košecká, J.J. Guerrero C. Sagüés, *Robotics and Autonomous Systems*, Vol. 56(6): 512-521 (Available online Available online 21 March 2008 - http://www.sciencedirect.com/).

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

Visual Door Detection Integrating Appearance and Shape Cues [⋆⋆]

A.C. Murillo [a,∗] J. Košecká [b] J.J. Guerrero [a] C. Sagüés [a]

[a] *DIIS - I3A, University of Zaragoza, Spain.*

[b] *Department of Computer Science, George Mason University, Fairfax, USA.*

Abstract

An 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 location 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 for door detection, by defining the likelihood of various features for generated door hypotheses. Differently 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 the structure of indoors environments. After the learning stage, we describe a hypothesis generation process and several approaches to evaluate the likelihood of the generated hypotheses. The approach is tested on numerous examples of indoor environments. It shows a good performance provided that the door extent in the images is sufficiently large and well supported by low level feature measurements.

Key words: door detection, generative models, geometry and appearance likelihood, indoor object recognition.

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

[∗] Corresponding author.
Email address: acm@unizar.es (A.C. Murillo).

1 Introduction

In this paper we present a new technique for detecting doors in perspective images of indoor environments using only visual information. Door detection is of great importance for various navigation and manipulation tasks. Doors are often places which separate different locations and can be used as landmarks for navigation and/or relative positioning or as waypoints to guide 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 in [4] and [5] doors are detected using both visual information and range data (sonar). In [4] authors exploit the fact that vision is good for providing long range information (beyond the range of ultrasound sensor). They detect and group vertical lines based on the expected door dimensions to form initial door hypotheses. In [5] the authors tackle the more general problem of obtaining a model of the environment. This model is defined by instantiations of several objects of predefined classes (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 the associated color values from the omnidirectional camera. In [6] both laser and camera data are integrated in such a manner that the trinocular vision sytem is used to select a possible door initial location and the laser measurements allow to dynamically update the door location while navigating towards it. In [7] authors focus on handling the variations in door appearance due to camera pose. They characterize 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 [8], 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 the door detection/recognition problem. Doors belong to a category of objects which do not have a very discriminative appearance, have quite discriminative shape and whose shape projection varies dramatically as a function of viewpoint, similarly to tables or shelves, for instance. The object recognition techniques explored extensively in computer vision commonly adopt the so called part based object models, which consider representations of objects in terms of parts [9] 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 a full generative model is sought or a discriminative technique is used, or combination of both. In the simplest of the generative model settings, the recognition stage proceeds with the computation of the posterior probability $P(Object|X, A)$, where X, A are the positions and appearance of the object parts. Assuming that the object can be characterized by a small number of parameters θ learned from training examples, the likelihood $P(X, A|\theta)$ can then be evaluated given the image measurements. With the discriminative approaches, multi-class classifiers are trained to distinguish between low-level features characteristic of a particular class [10] 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 [11]. There, the input features consist of geometric features computed from laser and Haar-like features computed from images.

Compared to part based representations [9], we pursue a model based approach. In this case, the geometry of the door is given and is specified by a small number of parameters and the appearance is learned from a few training examples. This type of representation resembles the models used in interpretation of architectural styles and man-made environments, where the analyzed scenes are typically well characterized by a small number of geometric/architectural primitives [12]. Instead of proposing the generative model of the whole image, we use the constraints of man-made environments to generate multiple hypotheses and use the learned probability distribution to evaluate their likelihood.

Outline

Section 2 describes the probabilistic model we adopt for door detection. We model the doors by a set of parameters, which are detailed together with their learnt models in section 3. The hypotheses generation process is explained in Section 4, followed in Section 5 by the likelihood evaluation process. Finally sections 6 and 7 present respectively door detection experiments and the conclusions of the work.

2 Problem formulation

We will assume that our door model 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 $P(Object|X, A)$, given the measurements X, A, which characterize the shape and appearance of the object hypotheses detected in the image:

$$P(Object|X, A) \propto P(X, A|Object)P(Object).$$

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., the 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 door width and height or the position of the bottom corners of the door, which should be touching the floor.

• $P(\theta_A)$ is the prior information on the appearance of the object, which 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 consider maximum likelihood values of the parameters θ_S and θ_A. For shape, they are are given by a known model and for appearance they are learned in a supervised setting. More details about these parameters will be provided in Section 3. The likelihood term can be further factored as follows, assuming that the appearance and the shape are independent attributes:

$$\begin{aligned} P(X, A|\theta_A, \theta_S) &= P(A|X, \theta_A, \theta_S)P(X|\theta_A, \theta_S) \qquad (2) \\ &= P(A|X, \theta_A)P(X|\theta_S). \end{aligned}$$

The shape likelihood evaluation is explained in Section 5.1 and the appearance likelihood evaluation approaches are described in Section 5.2.

3 Learning the model parameters

As mentioned before, our model for doors is described by a set of parameters characterizing its shape and appearance. While the shape parameters are given, the appearance parameters are learned from observations of labeled data. Later in the experimental section, some examples of the labeled images used for the learning process are shown (Fig. 11).

3.1 Shape representation

The shape model of our object of interest can be characterized by the parameters $\theta_S = [n_c, r]$, where $r = \frac{w}{h}$, being w and h the width and height of a door (see Fig. 1), and n_c is the number of corner features that support the model. Section 4.1 explains how these corner features are obtained from test images. These parameters characterize well the general shape of doors commonly encountered indoors.

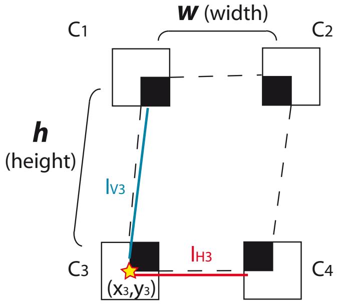


Fig. 1. Model of a door and components of one corner-feature fitted in the model (C_3): corner location (x_3, y_3) and lines (l_{H3}, l_{V3}) that give rise to the cross point.

The number of corners in the model is 4 and the ratio r is established to have a reference value λ_r in frontal views. This λ_r is 0.4 for typical observed samples.

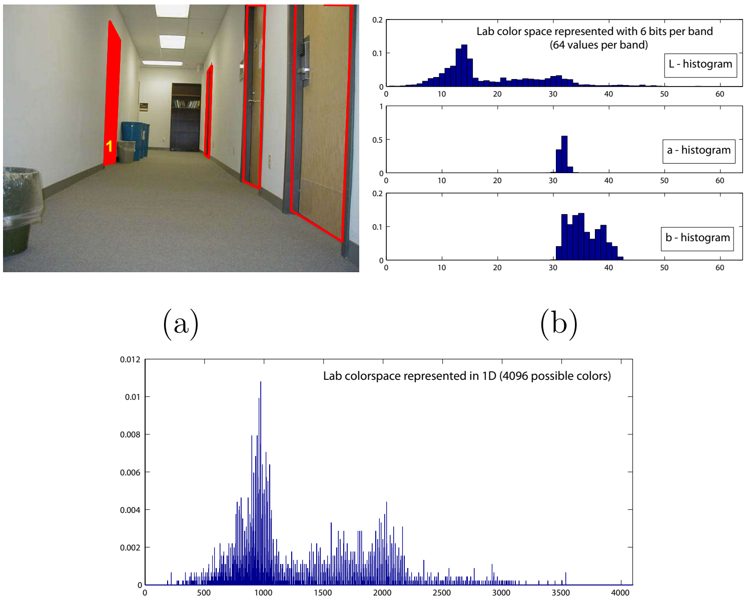
3.2 Appearance representation

The appearance parameters of the modelled object, θ_A, can be learned from the reference hand labeled door-segments shown in the experimental section. The main issue is how to represent the appearance information. Here we examine the most promising representations for the appearance among those studied in preliminary results on this work [13]. The appearance is represented by color signatures (histograms) computed for each door image region in the training set. We use the Lab (CIE 1976 (L* a* b*)) color space, which is commonly preferred since it best approximates the perceptually uniform color space [14].

Depending on the way the 3D color space is represented, three ways of building the histograms were considered:

- *Marginal* histograms with *fixed bin centers*. Each color band is quantized in n possible color values and is represented by a n-bins histogram. Therefore each region is represented by a $3n$ bin histogram. For instance, we quantize each color band to 6 bits (instead of the typical 8 bits representation, we use 6 as a compromise between accuracy and histograms dimension), then we have 2^6 possible values per color band. If we build a bin in the histogram for each possible value, a 64-bins histogram is computed for each of the three bands, as in the example in Fig. 2 (b). 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 [15] and has been shown to be useful in cases where there are few training examples (as it occurs in our case).
- *Full* histograms with *fixed bin centers*. In this case we consider the three color band values in a joint space. In our case the color space is quantized from the typical 24 bits representation (considering 256 possible values for each of the three color bands) to 12 bits (2^4 possible values for each band). In practice, each color band should be fitted to a range between 0 and 15, giving a set of 4096 ($2^4 \times 2^4 \times 2^4$) possible colors. Each 3D color $[L, a, b]$ now is represented in 1D with this value: $16^2 \frac{L}{255} 15 + 16 \frac{a}{255} 15 + \frac{b}{255} 15$. Including one bin in the histogram per each possible value, a 4096-bin histogram is computed for each region, as shown in the example in Fig. 2 (c).
- *Full* histograms with *variable bin centers*. As opposed to the two previous representations, in this case the bin centers in the histograms are variable and depend on the color distribution in each region. A clustering is performed on all pixel color values of each door region with a fixed number of clusters k. The color range comprised in the door examples is not too wide, therefore a small k value of 10 has been established experimentally as the most suitable. Each pixel in the region is assigned to the closest cluster and this constitutes the histogram, or more generally a signature of the region. Fig. 3 shows an example of the construction of this variable-bins histogram for the same door region shown in Fig. 2.

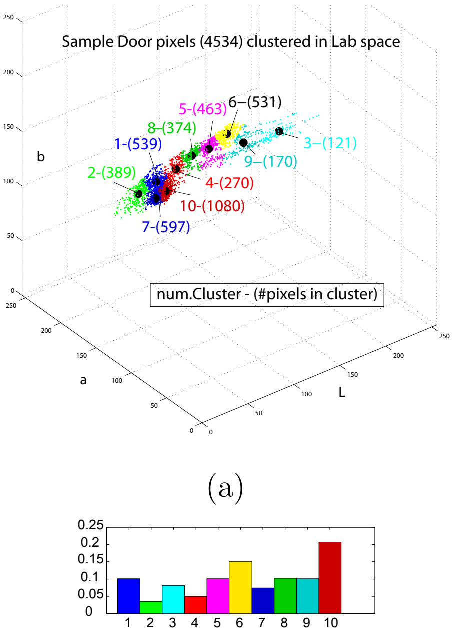
Once we have learned the models of the parameters describing the object of interest, we can evaluate the likelihood of all possible instantiations of the object model with regard to them.



(a) (b)

(c)

Fig. 2. (a) Sample door region 1 .(b) *Marginal* and (c) *Full* normalized histograms with fixed bin centers.



(a)

(b)

Fig. 3. (a) Door region pixels clustering: the centroid location of each cluster is marked with a thick black dot and the number of each cluster is plotted together with the number of pixels assigned to that cluster between parenthesis. (b) Normalized Variable bin center histogram.

4 Hypotheses generation

The selection of individual hypotheses on a certain image consists of two main steps: a low level features extraction process followed by the initial hypotheses instantiation.

4.1 Geometric primitives extraction

First, line segments are extracted from the image with our implementation of the approach described in [16] and the vanishing points are estimated following the approach described in [17]. Using the vanishing point information, 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 4 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 a square 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. 5).

4.2 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

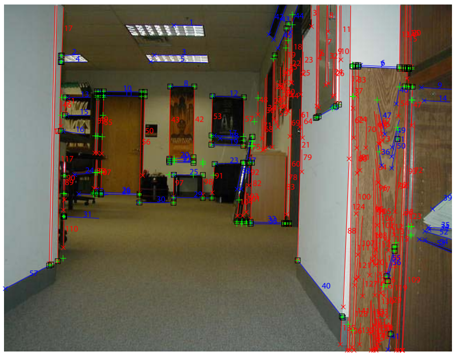


Fig. 4. Line segments grouped in vanishing directions (vertical in red, non-vertical in blue), corner points detected (green +) and corner points with high corner response (black □).

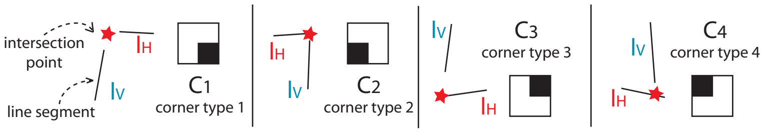


Fig. 5. Examples of line intersections considered and the four types of corner features that can be generated.

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. The search for two corner hypotheses is followed by the intersection between these sets of 2 compatible 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 $\{c_1, c_2, c_3, c_4\}$.

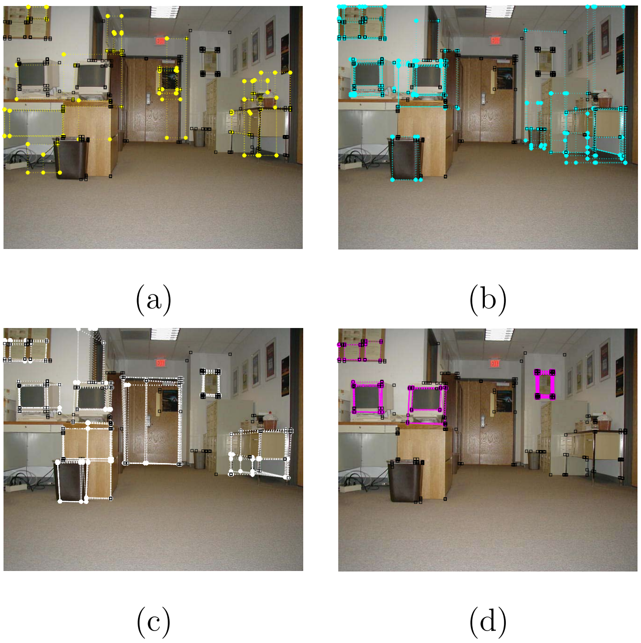

(a)
(b)
(c)
(d)

Fig. 6. Initial hypotheses generated for a test image. All extracted corner features are shown with a black square. (a) 1-corner hypotheses; (b) 2-corner hypotheses ; (c) 3-corner hypotheses; (d) 4-corner hypotheses.

After this stage, we have four types of hypotheses: supported by 4, 3 or 2 corner features or comprised by those singleton corners that did not have compatible corner features. Example hypotheses generated for an image are shown in Fig. 6. All extracted corner features are marked by a square (□), the corners contributing to each hypothesis are marked by $*$, and the dotted lines show the area delimited 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 line segments as shown in Fig. 7.

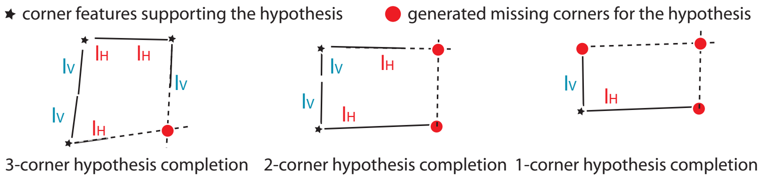

★ corner features supporting the hypothesis
● generated missing corners for the hypothesis
lV
lH
3-corner hypothesis completion
2-corner hypothesis completion
1-corner hypothesis completion

Fig. 7. Examples of the completion of initial hypotheses supported by less than 4 corner features.

5 Hypothesis evaluation - Likelihood computation

In this section we demonstrate the computation of the complete likelihood for each generated hypothesis being a door.

Let us suppose the image measurements X and A, for all generated hypotheses, which are characterizing some shape and appearance attributes respectively. In order to compute the probability $P(X, A|\theta)$, it can be decomposed as explained in Section 2 into a two factor expression: $P(X, A|\theta) = P(A|X, \theta_A)P(X|\theta_S)$.

5.1 *Evaluation of shape likelihood*

As described previously (section 3.1), the door model comprises two shape parameters $\theta_S = [n_c, r]$. Therefore, the likelihood of a hypothesis given these parameters is a combination of two terms

$$P(X|\theta_S) = P(X|n_c)P(X|r), \tag{3}$$

where X are the shape related measurements associated with the hypothesis, namely the supporting corners and the ratio of the associated region. The first term, $P(X|\theta_{n_c})$, was defined to assign higher likelihood to hypotheses which were supported by larger number of corner features. It consists of a discrete pdf defined as follows:

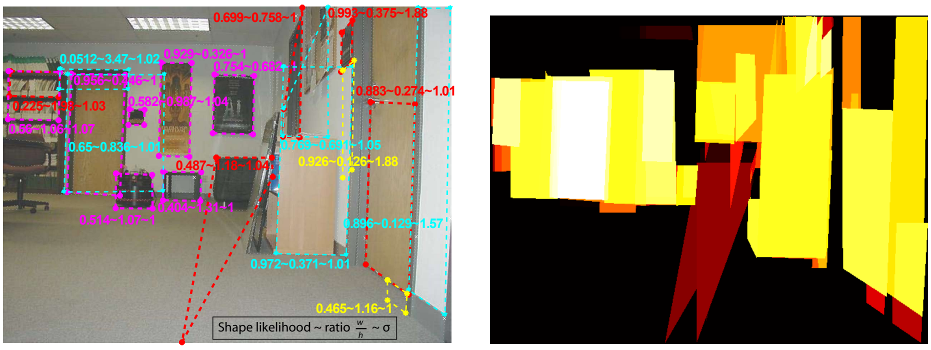
$$P(X|n_c) = 1 - 0.1(4 - n_c), \tag{4}$$

with $n_c \in [1, 2, 3, 4]$.

The second term $P(X|r)$ takes into account the ratio between the height and width of a door. We consider a typical ratio $\frac{w}{h}$ as $\lambda_r = 0.4$ for a common door in a frontal view in our observations. We evaluate the ratio between the height and width of the hypothesis and take into account how far is that hypothesis from a frontal view of a door. This is done by checking the perpendicularity of the lines composing the corner features. Then, the following shape likelihood terms is defined:

$$P(X|r) = e^{\frac{-(\frac{w}{h} - \lambda_r)}{\sigma^2}}, \tag{5}$$

with $\sigma^2 = \frac{\pi/2}{\alpha}$, where α is the angle, range $[0\ , \pi/2]$, between the non-vertical line and the vertical line that determined a certain corner. It takes into account the variability on the confidence of the ratio as a function of the viewpoint. In Fig. 8, there are several examples of hypotheses supported with different number of corners. For each hypothesis, the figure shows the probability obtained with eq. (5) followed by the ratio $\frac{w}{h}$ and the σ obtained as explained above. It can be seen there how the evaluation gracefully handles perspective distortion. For example, see the hypotheses on the right of the left image, they have the ratio $\frac{w}{h}$ far from the λ_r, but as they are identified as non frontal views, they are not as much penalized as other hypotheses also with bad ratio but frontal view (see hypotheses on the left of the same image). When a hypothesis is close to a fronto-parallel view, the ratio between width and height correctly penalizes "non-door" situations.



(a)

(b)

Fig. 8. Shape likelihood evaluation. (a) Examples of hypotheses with their $P(X|r)$, see eq. 5, and corresponding ratio $\frac{w}{h}$ and σ (for graphical simplification not all hypotheses details are plotted). (b) Example of shape likelihood of each hypothesis (the darker the color, the lower likelihood).

5.2 Evaluation of appearance likelihood

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, there are several ways how to represent the density function. This section focuses on the two main approaches followed to perform this evaluation. One is based on fixed bin centers histogram representation and the other one in the variable bin centers histogram representation.

- *Gaussian Mixture* based approach *(GM)*. This approach uses the fixed-bins location histograms explained previously (section 3.2). Both their full and marginal variants have been studied. The distribution of the reference (learned) color histograms is modelled as a mixture of Gaussians. The reference histograms, either marginal or full variations, are clustered with k-means and each cluster is represented by a centroid, mass and covariance. The probability of a certain region, obtained from the hypothesis generation, having door appearance depends on the distance, d_B, to the closest cluster:

  $$P(A|X, \theta_A) = e^{\frac{-d_B}{\sigma_h}}. \tag{6}$$

  The distance d_B between two normalized histograms h_1 and h_2 is based on the Bhattacharyya distance

  $$d_B(h_1, h_2) = 1 - \sum_{i=1}^{n} \sqrt{h_1(i)h_2(i)}, \tag{7}$$

  where n is the number of histogram bins and $h(i)$ is the weight of the i^{th} bin.
- *k-NN density estimation* based approach *(kNN)*. This approach uses the variable bin locations histograms explained previously (section 3.2). To evaluate the probability of a certain region being a door, we compute the histogram and the k-nearest neighbours among the learned door-region histograms are found. Then, the density estimation is performed with these k histograms as:

  $$P(A|X, \theta_A) = \frac{k}{nV}, \tag{8}$$

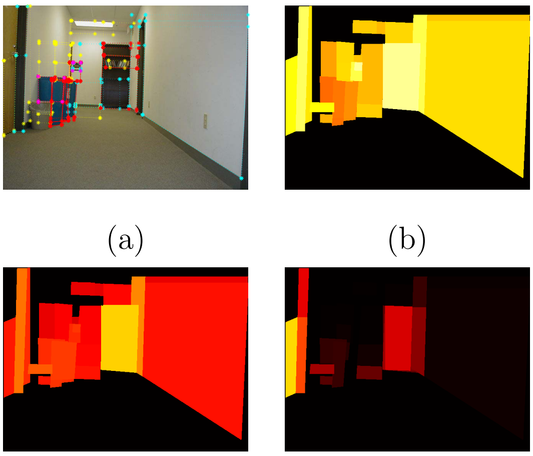
  being k the number of nearest neighbours selected from the n reference samples (learned histograms). V is the volume occupied by the k-neighbours.

  In this approach, the Earth Mover's distance (d_{EMD}) [18] is used to search for the closest histograms to a given one. It takes into account the bin center values and the weights of each bin, so it is very convenient for the variable bin representation used. This distance computed between two histograms $h_1 = \{(x_1, p_1), ..., (x_n, p_n)\}$ and $h_2 = \{(y_1, q_1), ..., (y_n, p_n)\}$ is obtained as:

  $$d_{EMD}(h_1, h_2) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}}, \tag{9}$$

  where x_i, y_i are the bin centroid values and p_i, q_i are the weights of each bin. d_{ij} is the ground distance between x_i and y_j and f_{ij} is the flow between x_i and y_j that minimizes the cost $\sum_{i=1}^{m} \sum_{j=1}^{n} d_{ij} f_{ij}$, as defined in [18].

Several likelihood-masks obtained for a particular example are shown in Fig. 9. They show the likelihood of the generated hypotheses using different approaches to represent the appearance parameters.



(a) (b)

(c) (d)

Fig. 9. Appearance likelihood evaluation masks. (a) Hypotheses generated in a test image. Evaluation with different approaches: (b) *GM*-marginal histograms, (c) *GM*-full histograms and (d) *k*-*NN*-full histograms.

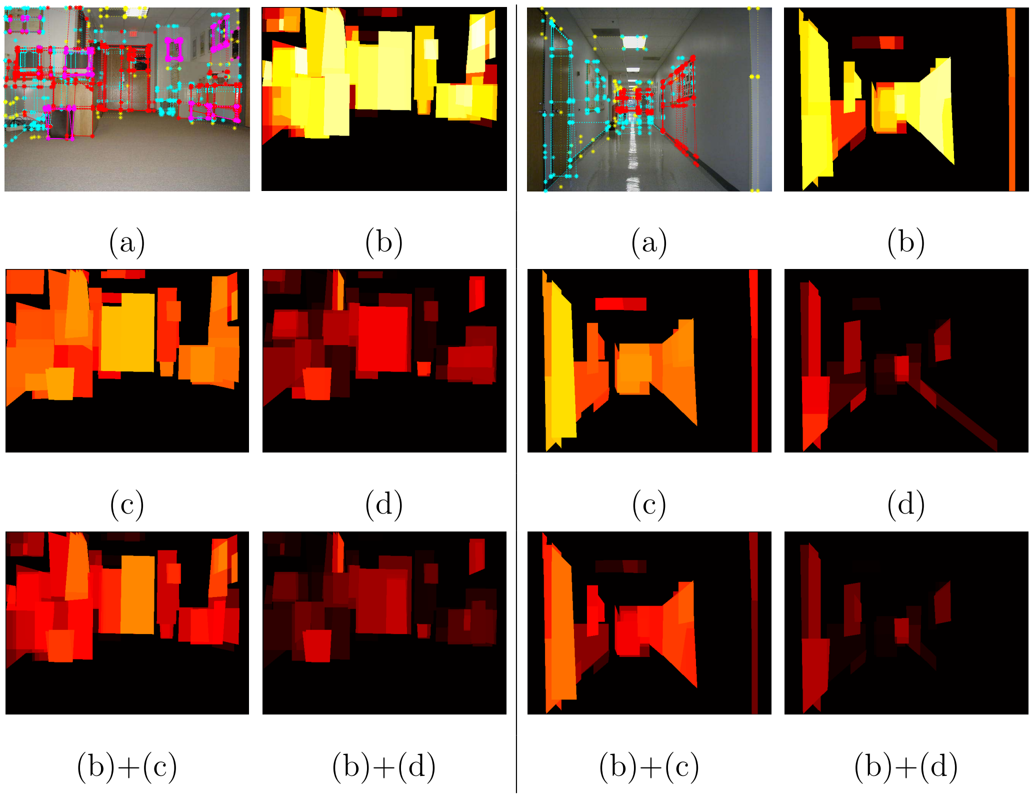
To obtain the complete likelihood evaluation, both shape and appearance terms should be integrated. Fig. 10 shows two examples with the obtained likelihood masks. There it is possible to observe the improvement when including both shape and appearance information, the door regions are better distinguished from the rest of the image in the likelihood masks that integrate both cues (last row of the figure).

6 Experimental Results

This section shows the most representative results from an extensive set of experiments, showing some advantages and disadvantages of the approaches studied. The experiments were performed with conventional images from different sets, some of them were acquired from our robots and hand-labeled (sets *GMU1* and *GMU2*), others were obtained from internet sources (set *web1*) to test the approach on more examples from different environments.

6.1 Reference image sets and labeling

The data sets *GMU1* and *GMU2* correspond to a robot tour around two indoor office-like environments. Only four images from each set were used for



(a) (b) (a) (b)

(c) (d) (c) (d)

(b)+(c) (b)+(d) (b)+(c) (b)+(d)

Fig. 10. Shape and appearance likelihood masks for two examples. (a) Test images with the generated hypotheses. (b) Shape likelihood masks. (c) Appearance likelihood masks using *GM* approach and (d) appearance likelihood masks using *k-NN* approach. The last row shows the integration of shape and each of the appearance evaluations. In all masks, the lighter the color the higher probability of being a door.

learning the appearance model. Images from one set contain 3 frontal views of doors and 3 oblique views of doors. The four images from the other set contain 9 oblique door views. All doors are wooden ones, except two of them that are elevator doors (metallic ones). These images were hand labeled by selecting the rectangular region covering approximately the door area in the image. Fig. 11 shows these reference images with their corresponding labeled doors. Around half of the remaining images from these data sets (37 images, with 76 doors), distributed along the two robot tours, were used to test the performance of the door recognition approaches.

The other images used to test the door detection were obtained from the web using image search engines (*web1*). The images were selected from a search done using the words: *door, wooden, corridor*. Fig. 12 shows some views from this set. These images were used only as test images to detect doors using the models learnt from our labeled data.

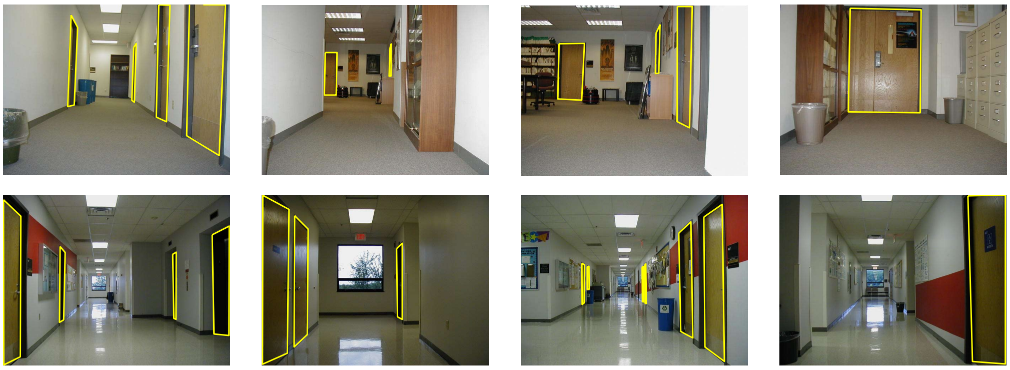
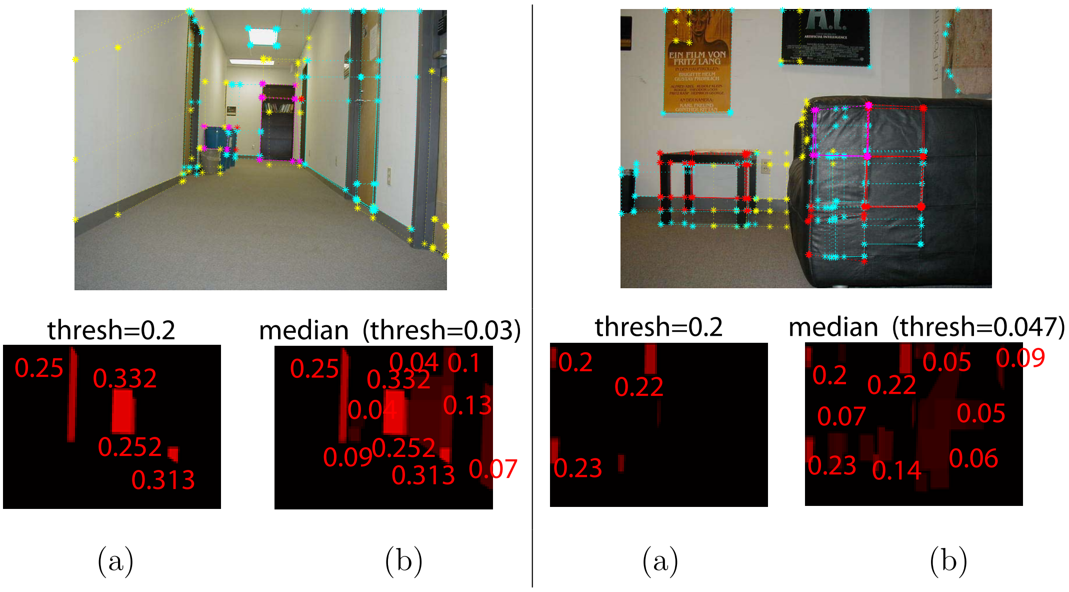


Fig. 11. Reference images from sets *GMU1* and *GMU2* used for the learning stage, with their corresponding hand-labeled door rectangular regions.

Fig. 12. Set *web1*: examples of test images obtained from the web.

6.2 Hypothesis acceptance criteria

In order to evaluate the door recognition performance, different criteria could be followed to decide which hypotheses are accepted as doors after the likelihood evaluation. As most hypotheses get a non zero probability of being a door, the most suitable approach is to establish a threshold to accept or reject hypotheses. Fixed and variable thresholds have been studied. The variable ones have shown better performance. They vary depending on the likelihood evaluation of the current image, e.g., the acceptance threshold is the median value of the likelihoods or a 75% of the maximum likelihood in the current test image. Fig. 13 shows two test images with the generated hypotheses in its top row. The bottom row of the same figure shows the hypotheses that were accepted for each example using two different criteria (thresholds): (a) with a fixed threshold of 0.2 and (b) with a variable threshold equal to the median value. The numbers in the images are the likelihood estimated for each accepted hypothesis. As we will confirm in next section with the more detailed performance evaluation (Table 1), the fixed thresholds handle better the cases without doors in the image (see right example in Fig. 13), however they seem to provide worse overall performance.



(a) (b) (a) (b)

Fig. 13. Hypothesis accepted using different approaches/criteria. Left: typical test image. Right: test image without doors.

6.3 Performance evaluation

Some detailed door detection performance results are shown in this section. They were obtained using the hand-labeled data sets (*GMU1*, *GMU2*), in order to have reference information to evaluate the recognition/detection rates.

Two different measures have been taken into account to quantitatively evaluate the recognition results. First, the number of doors detected from the doors actually visible in the images (% OK). Secondly, the % of pixels in the accepted hypotheses that were correctly or wrongly classified (% OK_{pix} or % FP_{pix} respectively). Fig. 14 shows several bar plots with these measurements for the different approaches using different acceptance thresholds.

Table 1 presents more detailed information about the approaches with better performance. Each row corresponds to a different approach. $\#doors$ is the total amount of doors in the test images, column OK shows the percentage of doors recognized and column OK_{pix} the percentage of all pixels from the reference labeled regions included in the accepted hypotheses. % OK is always higher or equal than % OK_{pix}, as accepted door hypotheses usually cover a smaller region in the image than the manually selected region for the evaluation. Columns $front$, $whole$ and $closed$ show the recognition rate if only the corresponding kind of door view is taken into account (frontal view, whole door view or closed door). As it could be expected, the easiest cases (i.e., frontal views and full door views) are almost always detected. Finally, column FP_{pix} contains the percentage of non-door pixels from the images included in the accepted hypotheses, and FP_{pix}^{noDoor} is the same measurement in the spe-

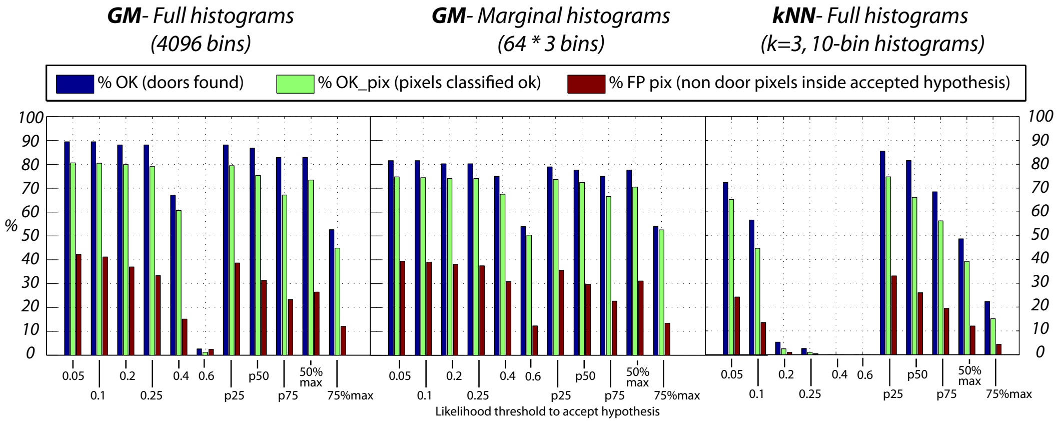


Fig. 14. Door recognition performance for the main approaches studied, with different acceptance criteria. 0.05,0.1,0.2,0.25,0.4,0.6 are fixed thresholds; pA (percentile A) and $B\%$max (B% of the maximum likelihood found) are variable thresholds.

cial case of images with no doors. All approaches gave a relatively high false positive rate, mainly due to test images with no doors, where the variable thresholds do not seem very suitable. As we can see from the results, fixed thresholds usually present lower amount of false positives.

Table 1
Door recognition evaluation: acceptance thresholds *percentile 25* (p25) and 0.2 or 0.05.

$\#doors$: 76	OK	$front$	$whole$	$closed$	OK_{pix}	FP_{pix}	FP_{pix}^{noDoor}
Full Hist - *GM* - *p25*	88%	100%	81%	88%	79%	39%	33%
Full Hist - *GM* - 0.25	88%	100%	81%	88%	79%	33%	32%
Marginal Hist -*GM* - *p25*	79%	92%	73%	81%	74%	36%	36%
Marginal Hist -*GM* - 0.25	80%	92%	76%	83%	74%	37%	24%
Full Hist - *k-NN* - *p25*	86%	92%	78%	85%	75%	33%	35%
Full Hist - *k-NN* - 0.05	72%	92%	73%	75%	65%	24%	18%

Some of the typically non-recognized doors are shown in Fig. 15. Note that most of them were too far, therefore they are too small to obtain reliable features, or they were highly occluded and no corner-features were extracted in the door region. Table 2 shows the results for the same experiments detailed in previous Table 1 when we do not take into account the doors with area smaller than a threshold (in this case with image area smaller than 2000 pixels, in test images of 640×480 pixels).

In these results, the main reason for failure in the recognition seems to be the lack of features for very small (distant) and narrow doors. This could be improved by adding alternative hypothesis instantiation possibilities where no cross points from lines were necessary. In mobile robot navigation, which is

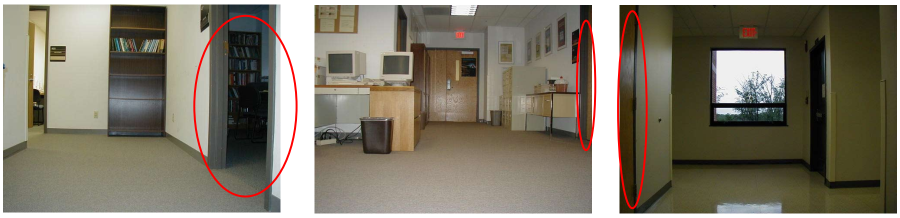


Fig. 15. Typical failure examples: the ellipses point some doors that are usually not correctly identified with any approach.

Table 2
Door recognition of doors over 2000 pixels in the image: acceptance threshold *percentile 25* (p25).

$\#doors$: 51	OK	$front$	$whole$	$closed$	OK_{pix}	FP_{pix}	FP_{pix}^{noDoor}
Full Hist - *GM*	90%	100%	82%	90%	79%	40%	33%
Marginal Hist - *GM*	84%	92%	75%	82%	75%	37%	32%
Full Hist - *k-NN*	86%	92%	79 %	85 %	72 %	33 %	34 %

our motivating application, the difficulty to recognize doors far away, due to their size, would decrease as the apparent size of the doors will increase as the robot moves around. The high percentage of false positives obtained with some approaches may be improved by adopting a more complex appearance representation or using additional information about the geometry of the environment, which would reject some of the hypotheses. This would be specially useful in environments with specular floors and walls, that show similar appearance than the searched object in the reflected area, for example see the image in Fig. 16.

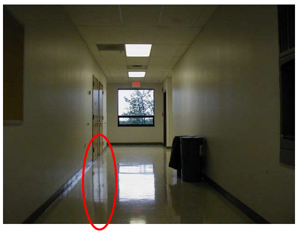
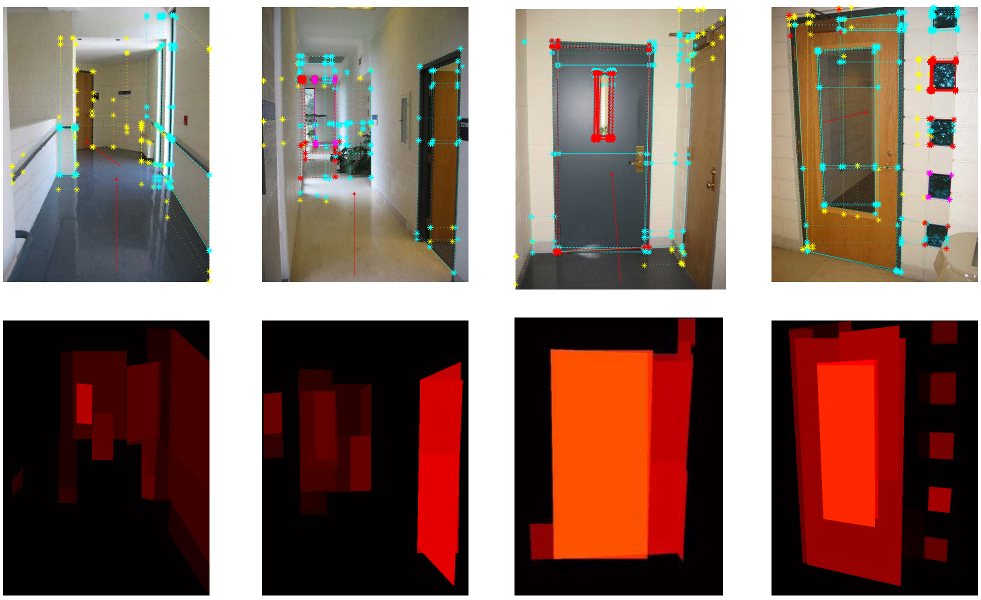


Fig. 16. Some reflectant-situations that cause many false positives.

6.3.1 Additional examples

In this section we present several examples of the "likelihood-masks" obtained for some of the web test images. As mentioned before, these test images come from different environments than the images used to learn the reference model. The good recognition results for some of the examples in Fig. 17 show that similar appearance (similar door materials) are enough to get a nice performance in door detection with our learned model. There are still many cases where it does not work properly. For instance the example (d) gives higher likelihood to the glass inner part of the door than to the wooden area. This is due to the configuration of the door, that is quite different from the doors used in the learning phase: this example is also wooden door but with a big glass area in the middle, however, any of the learned door region appearance signatures included a mixed distribution half wood half glass. On the other hand, the door detected in (c) correctly gets quite high likelihood even if it is not wooden door. Notice that the metal part has similar appearance than the metal frames from the doors in the learning stage. Since in the learning stage we had doors where mostly all the visible pixels were from the door frame, then metallic appearance regions had been also included in the appearance model.



(a) (b) (c) (d)

Fig. 17. Door detection in test images from the web. Top row: test images with the generated hypotheses. Bottom: hypotheses likelihood evaluation.

7 Conclusions

In this paper we have presented a new technique for detecting doors using only visual information. The probability distribution $P(Object|\theta)$ 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 a 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 door extent in the images is sufficiently large and well supported by low level feature measurements.

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 changes of the viewpoint, such as tables or shelves. Additional avenue which we would like to explore is to incorporate this technique in a robot mapping and exploration module in a dynamic setting. Alternative challenge not resolved by the proposed model is the capability of recognizing completely open doors, where the learned appearance is replaced by the clutter visible behind the door.

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] J. R. Asensio, J. M. M. Montiel, and L. Montano. Goal directed reactive robot navigation. In *IEEE Int. Conf. on Robotics and Automation*, pages 2905–2910, 1999.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] A. C. Murillo, J. Košecká, J. J. Guerrero, and C. Sagüés. Door detection in images integrating appearance and shape cues. In 2^{nd} *From Sensors to Human Spatial Concepts , held together with IROS 07*, pages 41–48, 2007.

[14] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[15] 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.

[16] 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.

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

[18] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *Int. Journal of Computer Vision*, 40(2):99–121, 2000.