2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Simultaneous Object Class and Pose Estimation for Mobile Robotic Applications with Minimalistic Recognition

Alper Aydemir, Adrian N. Bishop and Patric Jensfelt

Abstract— **In this paper we address the problem of simultaneous object class and pose estimation using nothing more than object class label measurements from a generic object classifier. We detail a method for designing a likelihood function over the robot configuration space. This function provides a likelihood measure of an object being of a certain class given that the robot (from some position) sees and recognizes an object as being of some (possibly different) class. Using this likelihood function in a recursive Bayesian framework allows us to achieve a kind of spatial averaging and determine the object pose (up to certain ambiguities to be made precise). We show how inter-class confusion from certain robot viewpoints can actually increase the ability to determine the object pose. Our approach is motivated by the idea of minimalistic sensing since we use only class label measurements albeit we attempt to estimate the object pose in addition to the class.**

I. Introduction

Object search (or active visual (object) search) is an important component of a mobile robot's action space [1]. For example, finding, identifying and localizing the pose of objects is a prerequisite for a robot that wishes to interact with objects in the environment. In [2] it is shown that a human understanding of space is significantly based on the objects present in the scene.

In this paper we are interested in the problem of simultaneous object class and pose estimation using a generic object classifier and a spatially dependent measurement likelihood model. One novelty we claim is the ability to estimate both the class and pose of objects in the environment given only measurements of the object class. Indeed, we attempt to push the limits of what information can be estimated given nothing more than a simple object class return and a model of the spatial likelihood for that class return.

Most existing classification and pose estimation algorithms match local geometric features *of the object*, such as corners, edges, holes and surfaces to a precise geometric model *of the object* [3]–[7]. Such techniques require extensive storage and training data and are far from minimalistic. In addition, these approaches are sensitive to object occlusions etc where object class measurements are possible but precise geometrical measurements of the object are not possible.

Other techniques [6], [8], [9] use a large number of labeled images taken from different poses and attempt to match specific images in order to determine the pose. The accuracy of these approaches increases with the amount of training and reference images. This technique critically ignores the relative geometry of the sensor and the object and the affect of this relationship on the likelihood of recognizing certain views (or more generally whole objects). In particular, we show how we can achieve more with much less input if we consider this relationship explicitly.

1) Original Contribution: We differ from existing object search methods in the design of our spatial likelihood functions. We will highlight throughout the paper that one novelty of our approach is that it is truly minimalistic in nature. By measuring only the object class label we attempt to extract both the true object class and object pose (orientation and location). Indeed, we generally ignore the notion of object view recognition and assign class labels only to entire objects. We certainly ignore any geometrical aspects of the object and we employ generic classifiers (we ignore the particular features employed by the classifier and indeed in our experiments we use a recognition algorithm from the literature which does not make use of any geometrical model of the object). We can extract certain estimates of the object pose purely from the structure of the likelihood functions which are defined over the robot configuration space and hence are geometrically related to the object pose. In addition, we show how inter-class confusion, e.g. the ability to mistakenly measure multiple class labels for a particular object type from certain views, can be advantageous to the estimation problem (specifically the estimation of object pose). We can achieve a high degree of accuracy in pose estimation using our technique and exploiting inter-class confusion. As far as we are aware our technique is novel and truly minimalistic.

2) Paper Outline: This paper is organized as follows. In the next section we outline the general notation used throughout the paper along with the basics of the robot dynamic model considered. In Section III we formulate the problem and outline the design of the likelihood function. We also provide some intuition regarding the design of the spatial likelihood function through example. Furthermore, in Section III we outline the recursive Bayesian algorithm for computing an objects pose and class and we highlight the algorithm behaviour using a simple toy example. We show how measurements of the class label alone can be used to determine accurately the object pose given a suitable likelihood function defined over the robot configuration space. We then outline an extension of the algorithm in Section IV for object class and orientation estimation over a grid. In Section V we provide the results of a practical experiment over a grid and in Section VI we discuss the results and directions for future work. Our conclusion is given in Section VII.

The authors are with the Centre for Autonomous Systems (CAS) at the Royal Institute of Technology (KTH), Stockholm, Sweden. This work was supported by the Swedish Foundation for Strategic Research (SSF) through CAS and also via the EU FP7 project CogX.

978-1-4244-5040-4/10/$26.00 ©2010 IEEE

II. Preliminaries

In this section we outline some notational preliminaries and the robot dynamical model considered.

A. Notation

Introduce a global coordinate frame $\mathcal{C}$ at some pre-defined time t_0. Consider a set of *objects* $\mathcal{O} = \{o_1, \ldots, o_{n_o}\}$ with position $\mathbf{x}_i \in \mathbb{R}^2$ and orientation $\phi_i \in \mathrm{S}^1$. Consider an arbitrary object o_i placed so that the $\mathbb{R}^2$ location of the object's center hovers over the origin of $\mathcal{C}$ at t_0. Introduce a local two-dimensional coordinate frame $\mathcal{C}_i$ at the center of o_i. Then the orientation ϕ_i is defined as the relative rotation of $\mathcal{C}_i$ with respect to $\mathcal{C}$. Each object o_i belongs to a class $\{c_j\}_{j=1}^{n_c}$ or the *unclassified* or *non-object* class c_0.

The position of a single mobile robot is denoted by $\mathbf{s} \in \mathbb{R}^2$ with heading $\theta \in \mathrm{S}^1$. The distance between the robot and o_i is given by $r_i = \|\mathbf{x}_i - \mathbf{s}\|$. The relative direction to o_i from the robot's heading is given by $\vartheta_i = \alpha_i - \theta$ where α_i is the azimuthal bearing to o_i in the global coordinate system and $\vartheta_i \in \mathrm{S}^1$. We then define a viewpoint $\mathbf{p}_i = [\mathbf{s}\ \vartheta_i]^\top$.

B. Dynamics

Introduce the matrix Lie group $\mathbb{SE}(2)$ with group element $\mathbf{X}(\psi, \mathbf{q}) \in \mathbb{SE}(2)$ with $\mathbf{q} = [x\ y]^\top \in \mathbb{R}^2$ and a group (matrix) multiplication operator. An element $\mathbf{X}(\theta, \mathbf{r}) \in \mathbb{SE}(2)$ acts on a point $\mathbf{p}_i \in \mathbb{R}^2$ by mapping it to $(\mathbf{R}(\theta)\mathbf{p}_i + \mathbf{r}) \in \mathbb{R}^2$. Here $(\mathbf{R}(\theta)$ is the rotation matrix defined as

$$\mathbf{R}(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \tag{1}$$

and note that all elements in $\mathbb{SO}(2)$ are congruent to such a matrix. For notational brevity we write the action of $\mathbf{X}(\psi, \mathbf{r})$ on $\mathbf{q}$ as

$$\mathbf{X}(\psi, \mathbf{r}) \circ \mathbf{q} = \mathbf{R}(\psi)\mathbf{q} + \mathbf{r} \tag{2}$$

which constitutes a left action of $\mathbb{SE}(2)$ on $\mathbb{R}^2$. The inverse $\mathbf{X}^{-1}(\theta, \mathbf{r}) \in \mathbb{SE}(2)$ maps $\mathbf{p}_i$ to $\mathbf{R}^\top(\theta)\mathbf{p}_i - \mathbf{R}^\top(\theta)\mathbf{r}$ and the identity is given by $\mathbf{X}(0, \mathbf{0}) \in \mathbb{SE}(2)$.

Associated with $\mathbb{SE}(2)$ is the vector space $\mathrm{se}(2)$ which is a Lie algebra with respect to the Lie bracket operation. We define the basis of $\mathrm{se}(2)$ by $\{\mathbf{E}_x, \mathbf{E}_y, \mathbf{E}_\psi\}$ with

$$\mathbf{E}_i = \begin{bmatrix} 0 & 0 & 1(i=x) \\ 0 & 0 & 1(i=y) \\ 0 & 0 & 0 \end{bmatrix}, \ \mathbf{E}_\theta = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3}$$

with $i \in \{x, y\}$ and where $1(\cdot)$ is an indicator function.

Given translational and angular velocity control inputs $u_1 = v$ and $u_2 = \omega$ we then have

$$\dot{\mathbf{X}}(\theta, \mathbf{t}) = \mathbf{X}(\theta, \mathbf{t})\left(\mathbf{E}_x u_1 + \mathbf{E}_\theta u_2\right) \tag{4}$$

which constitutes a left-invariant, drift-free system on the group $\mathbb{SE}(2)$. This model is the Lie group representation of the unicycle model and is our robot kinematic model.

III. Problem Formulation

In this section we outline the probabilistic framework within which our estimation problem is formulated.

A. Classification Likelihoods on Lie Groups

For each $\mathbf{p}(t)$ the robot takes measurements of the potential class of o_i in the form

$$\mathbf{y}_i(t) = [\widehat{c}_j\ \ldots\ \widehat{c}_k]^\top \tag{5}$$

with $\mathbf{y} = [\mathbf{y}_1^\top\ \ldots\ \mathbf{y}_{n_o}^\top]^\top$. This means that a measurement of object o_i can return more than a single class value[1].

We model the likelihood of measuring $\widehat{c}_j$ for o_i as a function of the robot pose. In fact, we model this likelihood as a sum of Gaussian densities on the Lie group $\mathbb{SE}(2)$. Consider an arbitrary normal density in $\mathbb{R}^n$ of the form

$$\gamma(\mathbf{x} - \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(\frac{1}{2}\|\boldsymbol{\Sigma}^{\frac{-1}{2}}(\mathbf{x} - \boldsymbol{\mu})\|_2^2\right) \tag{6}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix and $\boldsymbol{\mu}$ is the mean. A Gaussian distribution on the Lie group $\mathbb{SO}(2)$ is given by

$$\chi(x - \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k \in \mathbb{Z}} \exp\left[\frac{-(x - \mu - 2\pi k)^2}{2\sigma^2}\right] \tag{7}$$

and if $\sigma^2 << 2\pi$ in $\chi(x - \mu, \sigma^2)$ then $\chi(x - \mu, \sigma^2)$ can be approximated well by the case $k = 0$ in (7). A Gaussian on the product space $\mathbb{SE}(2)$ can then be denoted by $\zeta(\mathbf{x} - \boldsymbol{\mu}, \boldsymbol{\Sigma})$. We state the following lemma for completeness.

Lemma 1 ([10]): There exists an integer m and constants $w_i > 0$ with $\sum_{i=1}^m w_i = 1$, such that the Gaussian sum

$$p_{approx}(\mathbf{x}) = \sum_{i=1}^{m} w_i \gamma(\mathbf{x} - \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \tag{8}$$

can approximate any density function $p(\mathbf{x})$ as closely as desired in the sense that $\int_{\mathbb{R}^n} |p(\mathbf{x}) - p_{approx}(\mathbf{x})|\ d\mathbf{x}$ can be made arbitrarily small.

Recall that the element $\mathbf{X}(\psi, \mathbf{r}) : \mathbb{R}^2 \to \mathbb{R}^2$ acts on points via left translation denoted by $\mathbf{X}(\psi, \mathbf{r}) \circ \mathbf{q}$. For notational brevity we introduce the following notational definition

$$\mathbf{X}(\psi, \mathbf{r}) \circ [\mathbf{q}\ q_3\ q_4\ \ldots\ q_n]^\top = [\mathbf{R}(\psi)\mathbf{q} + \mathbf{r}\ q_3\ q_4\ \ldots\ q_n]^\top \tag{9}$$

which means that $\mathbf{X}(\psi, \mathbf{r}) : \mathbb{R}^2 \times \mathcal{A} \to \mathbb{R}^2 \times \mathcal{A}$ by acting on the first two dimensions in the standard way and leaving the remaining $n - 2$ dimensions unchanged.

We model the likelihood function by

$$p(\widehat{c}_i, o_j | c_i, \phi_j, \mathbf{x}_j) = \mathrm{P}(\widehat{c}_i, o_j | c_i) \sum_{k_i} \frac{w_{k_i}}{(2\pi)^{\frac{3}{2}}|\boldsymbol{\Sigma}_{k_i}|^{1/2}} \times \exp\left(\frac{1}{2}\|\boldsymbol{\Sigma}_{k_i}^{\frac{-1}{2}}(\mathbf{p}_j - \mathbf{X}(\phi_j, \mathbf{x}_j) \circ \mathbf{q}_{k_i})\|_2^2\right) \tag{10}$$

where $\sum_{k_i} w_{k_i} = 1$ with $i = \{1, \ldots, n_c\}$. For an object with position $\mathbf{x}_j$ and orientation ϕ_j we define $\mathbf{X}(\phi_j, \mathbf{x}_j)\mathbf{q}_{k_i}$

[1] For example, observing a car from the front may yield several positive car model class returns). We assume perfect data association, i.e. we know which objects o_i generate particular class measurements.

with $\mathbf{X}(\phi_j, \mathbf{x}_j) \in \mathbb{SE}(2)$ and $\mathbf{q}_{k_i} \in \mathbb{R}^2 \times \mathbb{SO}(2)$ as the *mean* of the ${k_i}^{th}$ Gaussian and $\mathbf{\Sigma}_{k_i}$ is the *covariance*[2].

The term $\mathrm{P}(\widehat{c}_i, o_j | c_i)$ specifically deals with the likelihood of o_j being of class c_i given the measurement $\widehat{c}_i$ whereas $p(\widehat{c}_i, o_j | c_i, \phi_j, \mathbf{x}_j)/\mathrm{P}(\widehat{c}_i, o_j | c_i)$ is the likelihood of o_j being in position $\mathbf{x}_j$ with orientation ϕ_j.

The likelihood $p(\widehat{c}_i, o_j | c_i, \phi_j, \mathbf{x}_j)$ is also a probability density function such that for bounded regions of $\mathcal{A}$ of $\mathbb{SE}(2)$ with positive Lebesgue measure the integral

$$\int_{\mathcal{A}} p(\widehat{c}_i, o_j | c_i, \phi_j, \mathbf{x}_j) \; d\mathbf{p} \tag{11}$$

gives the probability of measuring $\widehat{c}_i$ for o_j given that o_k is of class c_i and with orientation ϕ_j and position $\mathbf{x}_j$. We state this explicitly since we will require that the so-called *confusion densities* $p(\widehat{c}_j, o_k | c_i, \phi_k, \mathbf{x}_k)$ with $i \neq j$ satisfy the inequality

$$\int_{\mathcal{A}} p(\widehat{c}_j, o_k | c_i, \phi_k, \mathbf{x}_k) \; d\mathbf{p} \leq \int_{\mathcal{A}} p(\widehat{c}_i, o_k | c_i, \phi_k, \mathbf{x}_k) \; d\mathbf{p} \tag{12}$$

or $p(\widehat{c}_j, o_k | c_i, \phi_k, \mathbf{x}_k) \leq p(\widehat{c}_i, o_k | c_i, \phi_k, \mathbf{x}_k)$ for all bounded subsets $\mathcal{A}$ of $\mathbb{SE}(2)$ in a defined region of interest $\mathcal{R} \subset \mathbb{SE}(2)$. That is, over any bounded region in $\mathcal{R}$ we want the probability of measuring $\widehat{c}_j$ for o_i to be less than (or equal to) the probability of measuring c_i given that the object is of true class c_i (and for all object poses).

Of course, this inequality cannot be satisfied over all of $\mathbb{SE}(2)$ if the confusion likelihood is also required to be a true density function. But to make this definition consistent we note that when viewed as a likelihood function $p(\widehat{c}_i, o_j | c_k, \phi_j, \mathbf{x}_j)$ is valid as long as it is congruent to a probability density function via multiplication by a constant.

We model the likelihood function of false positives by the following Gaussian mixture

$$p(\widehat{c}_i, o_j | c_k, \phi_j, \mathbf{x}_j) = \textstyle\sum_{k_{ik}} \frac{\mathtt{common}(k_i, k_k) \;\; w_{k_{ik}}}{(2\pi)^{\frac{3}{2}} |\mathbf{\Sigma}_{k_{ik}}|^{1/2}} \times$$
$$\exp\left(\tfrac{1}{2} \| \mathbf{\Sigma}_{k_{ik}}^{\frac{-1}{2}} \left(\mathbf{p} - \mathbf{X}(\phi_j, \mathbf{x}_j) \circ \mathbf{q}_{k_{ik}}\right) \|_2^2\right) \mathrm{P}(\widehat{c}_i, o_j | c_k) \tag{13}$$

where the sum over k_{ik} has at most $\min(k_i, k_k)$ terms and (12) must hold in $\mathcal{R} \subset \mathbb{SE}(2)$. The function

$$\mathtt{common}(k_i, k_k) \in \{0, 1\} \tag{14}$$

captures the fact that an object o_j can be confusingly observed as $\widehat{c}_i$ and/or $\widehat{c}_k$ from some robot positions because the underlying true classes c_i and c_k share a *common* indistinguishability from such locations[3].

The class c_0 is used to model unclassified classes or locations in space where no object exists. The likelihood $p(\widehat{c}_i, o_j | c_0, \phi_j, \mathbf{x}_j)$ where $i \neq 0$ is given by

$$p(\widehat{c}_i, o_j | c_0, \phi_j, \mathbf{x}_j) = \mathrm{P}(\widehat{c}_i, o_j | c_0) \tag{15}$$

which although not a true likelihood function is valid over any bounded region $\mathcal{R} \subset \mathbb{SE}(2)$ since it is congruent to a uniform density over $\mathcal{R}$. For all classes for which it is defined we now require $\sum_i \mathrm{P}(\widehat{c}_i, o_j | c_k) = 1$.

For much of the space $\mathbb{SE}(2)$ the object recognizer will not return any class value for o_j. We can (if desired) model the absence of any returns in $\{c_1, \ldots, c_{n_c}\}$ as a measurement of the dummy class c_0. We would then need to construct the likelihood $p(\widehat{c}_0, o_j | c_i, \phi_j, \mathbf{x}_j)$. We do not explore the design of this likelihood in detail since we will not (in our implementations) incorporate dummy measurements when no class is detected[4].

If we define $\mathbf{c} = [c_0 \; c_1 \; \ldots \; c_{n_c}]$ then the likelihood

$$p(\widehat{c}_i, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) = \sum_k p(\widehat{c}_i, o_j | c_k, \phi_j, \mathbf{x}_j) \tag{16}$$

is the multi-dimensional likelihood function of the object being in all of the defined classes and all poses given a particular class return. Given a return measurement $\mathbf{y}_i(t) = [\widehat{c}_a \; \widehat{c}_b \; \ldots \; \widehat{c}_z]^\top$ for object i then the joint likelihood is

$$p(\mathbf{y}_j, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) = \prod_{\widehat{c}_k \in \mathbf{y}_j} p(\widehat{c}_k, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) \tag{17}$$

under a naive Bayesian assumption, i.e. under the assumption that $p(\widehat{c}_k, \cdot | \mathbf{c}, \cdot, \widehat{c}_j) = p(\widehat{c}_k, \cdot | \mathbf{c}, \cdot)$.

B. Example Likelihood Functions

We now provide some intuition regarding the design of the likelihood functions. These examples are simplified but illustrate the heuristics behind the likelihood structure.

Consider an object o_1 of class c_1 located at the origin at time t_0 with defined orientation $\phi_1 = 0$. An object classifier is trained on object o_1 from a number of relative positions denoted by $\mathbf{q}_{k_1}$ with $\mathbf{q}_{k_1} = [q_{k_1}^1 \; q_{k_1}^2 \; 0]^\top$. For the k^{th} training position we define a Gaussian $\zeta(\mathbf{p}_1 - \mathbf{X}(\phi_1, \mathbf{x}_1) \circ \mathbf{q}_{k_1}, \mathbf{\Sigma}_{k_1})$ where $\mathbf{\Sigma}_{k_1}$ is tuned based on the specific classifiers properties[5]. We define $p(\widehat{c}_1, o_i | c_1, \phi_i, \mathbf{x}_i)$ as the sum of such Gaussians as in (10) with $w_{k_1} = 1/4$ and $\mathrm{P}(\widehat{c}_1, o_1 | c_1) = 1$. In this example we set $\mathbf{q}_{1_1} = [10 \; 0 \; 0]^\top$, $\mathbf{q}_{2_1} = [-10 \; 0 \; 0]^\top$, $\mathbf{q}_{3_1} = [0 \; 10 \; 0]^\top$ and $\mathbf{q}_{4_1} = [0 \; -10 \; 0]^\top$ with $\mathbf{\Sigma}_{k_1} = \mathrm{diag}(10, 10, \pi/4)$. Now consider a random object o_i at $\mathbf{x}_i =$

[2] We note at this point that the Gaussian parameters $\mathbf{q}_{k_i}$ and $\mathbf{\Sigma}_{k_i}$ are defined based on the training scheme of the object classifier and $\mathbf{p}_j$ is the robot position when the relevant class labels are measured. Heuristically, $\mathbf{q}_{k_i}$ is taken to be (one of) the sensor's position in $\mathbb{SE}(2)$ at training time relative to the object (which is located at the origin during training with the reference orientation). The variance is (in this paper) tuned to provide a realistic model of the spatial dependence of the recognition algorithm at run time to the trained classifier models. In the next subsection we provide an example further illustrating how the likelihood functions are created. However, we note here that the motivation for these likelihood functions is motivated from experience where we have noticed that often simply by measuring the class label for an entire object (not view point) we *most likely* restricted one of a small number of points. In reverse, given a known robot position, the object is *most likely* in one of a small number of locations with one of a small number of orientations.

[3] Such a case happens, for example, if there are object types which look very similar or share a similar internal representation (ambiguous objects or object-data) from certain views.

[4] The reason we do not generate dummy measurements is that $\widehat{c}_0$ will, in general, provide little information about the true class c_j (including c_0) for most robot positions $\mathbf{p}$ and objects o_i. Heuristically, over bounded regions the likelihood $p(\widehat{c}_0, o_j | c_i, \phi_j, \mathbf{x}_j)$ would resemble a constant minus the sum (16). Of course this would require some further justification in order for $p(\widehat{c}_0, o_j | c_i, \phi_j, \mathbf{x}_j)$ to be valid as a likelihood function.

[5] We discuss later that an interesting direction for future work is the design of reinforcement learning schemes for tuning such parameters.

$[5\ 5]^\top$ with $\phi_i = 45^o$. We plot $p(\widehat{c}_1, o_i | c_1, \phi_i, \mathbf{x}_i)$ with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$ in Figure 1.

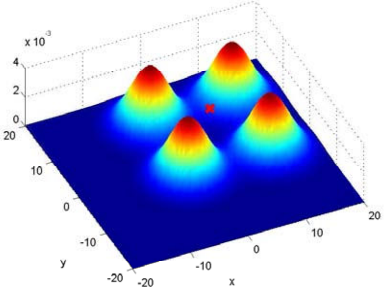


Fig. 1. An example likelihood function with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$.

Figure 1 shows from which positions in space relative to the target object o_i the likelihood of the class of o_i being c_1 is given that we measure $\widehat{c}_1$ (and in essence assume robot relative heading ϑ_i invariance - e.g. this will hold for omni-directional cameras). Note that the likelihood just demonstrated is symmetric in terms of the orientation ϕ_i, and as a result ϕ_i can be determined only up to rotations modulo $\pi/2$ given measurements of the class $\widehat{c}_1$. We outline the specific estimation technique in the next subsection and later provide examples of the pose accuracy that can be achieved.

However, we now demonstrate how inter-class confusion can be aid the pose estimation problem if the views from which the confusion is likely are not maximal. Consider an object o_2 of class c_2 located at the origin at time t_0 with defined pose $\phi_2 = 0$. An object classifier is trained on object o_2 from a number of relative positions denoted by $\mathbf{q}_{k_2}$ and for the k^{th} position we define a Gaussian $\zeta(\mathbf{p}_2 - \mathbf{X}(\phi_2, \mathbf{x}_2) \circ \mathbf{q}_{k_2}, \mathbf{\Sigma}_{k_2})$. Then we define $p(\widehat{c}_2, o_i | c_2, \phi_i, \mathbf{x}_i)$ as the sum of such Gaussians as in (10) with $w_{k_2} = 1/4$ and $\mathrm{P}(\widehat{c}_2 | c_2) = 1$. In this example we set $\mathbf{q}_{i_2} = \mathbf{q}_{i_1}$ and $\mathbf{\Sigma}_{k_2} = \mathbf{\Sigma}_{k_1}$. Now consider a random object o_i at $\mathbf{x}_i = [5\ 5]^\top$ with $\phi_i = 45^o$. The plot of $p(\widehat{c}_2, o_i | c_2, \phi_i, \mathbf{x}_i)$ with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$ is identical in this case to the likelihood shown in Figure 1.

Now suppose from a number of positions we know that o_1 and o_2 can be confusingly recognized as both c_1 and c_2 in some instances. In this example,

$$\mathfrak{common}(2_1, 2_2) = 1 \tag{18}$$

and all other $\mathfrak{common}(\cdot)$ equal to zero. We let $\mathbf{\Sigma}_{k_{1,2}} = \mathbf{\Sigma}_{k_2} = \mathbf{\Sigma}_{k_1}$ and $w_{k_{1,2}} = 1/4$. Also let $\mathrm{P}(\widehat{c}_2 | c_1) = \mathrm{P}(\widehat{c}_1 | c_2) = 1/2$ and now let $\mathrm{P}(\widehat{c}_i | c_i) = 1/2$. Then $p(\widehat{c}_j, o_i | c_j, \phi_i, \mathbf{x}_i)$, for $j = \{1, 2\}$, with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$ is the same shape as in Figure 1 except both likelihoods are weighted by $1/2$. In Figure 2 we plot $p(\widehat{c}_1, o_i | c_2, \phi_i, \mathbf{x}_i) = p(\widehat{c}_2, o_i | c_1, \phi_i, \mathbf{x}_i)$ with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$ for the same random object o_i at $\mathbf{x}_i = [5\ 5]^\top$ with $\phi_i = 45^o$.

Given a random object o_i of class c_1 or c_2 we now gain some intuition about how class confusion can aid in removing any ambiguity regarding the pose of the object. For example, if o_i were viewed from a number of robot positions around $\mathbf{X}(\phi_i, \mathbf{x}_i) \circ \mathbf{q}_{2_1}$, i.e. around the confusion peak, and both $\widehat{c}_1$ and $\widehat{c}_2$ were measured at these positions then the likelihood of the object pose would be (significantly) dominated by a single mode at the true pose. We will explore a detailed toy example illustrating this property later. We will also explore a practical example showing a real-world experimental result.

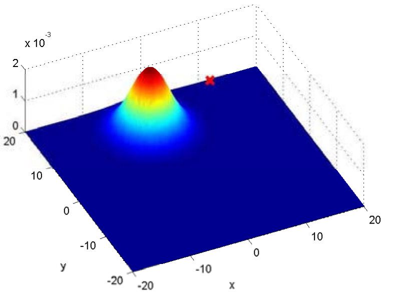


Fig. 2. A confusion likelihood function with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$.

C. Maximum A Posterior Probabilities

In terms of Bayes' rule we know

$$p(c_i, \phi_j, \mathbf{x}_j | \widehat{c}_i, o_j) = \frac{p(\widehat{c}_i, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) p(c_i, \phi_j, \mathbf{x}_j | o_j)}{p(\widehat{c}_i, o_j)} \tag{19}$$

or in terms of $\mathbf{y}(t)$ we have

$$p(c_i, \phi_j, \mathbf{x}_j | \mathbf{y}_j, o_j) = \frac{p(\mathbf{y}_j, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) p(c_i, \phi_j, \mathbf{x}_j | o_j)}{p(\mathbf{y}_j, o_j)} \tag{20}$$

where the denominator is given by

$$p(\mathbf{y}_j, o_j) = \int_{\mathbb{SE}(2)} p(\mathbf{y}_j, o_j | \mathbf{c}, \phi_j, \mathbf{x}_j) p(c_i, \phi_j, \mathbf{x}_j | o_j)\ d\phi_j d\mathbf{x}_j \tag{21}$$

Note we have neglected illustrating the dependence on time but the recursion is clear with the prior $p(c_i, \phi_j, \mathbf{x}_j)$ at time t equal to the posterior $p(c_i, \phi_j, \mathbf{x}_j | \mathbf{y}_j, o_j)$ computed at some time $\tau < t$. We also know that

$$\mathrm{P}(c_i | \mathbf{y}_j, o_j) = \int_{\mathbb{SE}(2)} p(c_i, \phi_j, \mathbf{x}_j | \mathbf{y}_j, o_j)\ d\phi_j d\mathbf{x}_j \tag{22}$$

is the posterior probability of object o_j being of class c_i given the measurements $\mathbf{y}_j$ and $\sum_i \mathrm{P}(c_i | \mathbf{y}_j, o_j) = 1$ where $i = 0$ can be included naturally. For any o_j we then have

$$\sum_i \int_{\mathbb{SE}(2)} p(c_i, \phi_j, \mathbf{x}_j | \mathbf{y}_j, o_j)\ d\phi_j d\mathbf{x}_j = 1 \tag{23}$$

where the sum is taken over all classes c_0 to c_{n_c}.

If we want the maximum a posterior (MAP) class and object orientation (or pose) then we can take the maximum class index and object pose estimates via

$$\{\widetilde{c}_i, \widetilde{\phi}_j, \widetilde{\mathbf{x}}_j\} = \underset{i, \phi_j, \mathbf{x}_j}{\operatorname{argmax}} \{p(c_i, \phi_j, \mathbf{x}_j | \mathbf{y}_j, o_j)\}_{i \in \{1, \ldots, n_c\}} \tag{24}$$

where $\widetilde{c}_i$ is the MAP class estimate for object j corresponding to the maximization argument index i.

In general, (24) leads to n_c maximization problems for each o_j. Each density is often multi-modal but each mode can be determined easy via grid-search. If $\widehat{c}_k \notin \mathbf{y}_j(t)$ and $p(\widehat{c}_k, o_j|c_i, \phi_j, \mathbf{x}_j) = 0$ for all $i \neq k$ then $p(c_i, \phi_j, \mathbf{x}_j|\mathbf{y}_j, o_j) = 0$ at time t and at least one maximization problem is avoided.

D. Bringing it All Together with a Toy Example

A toy example is now examined in order to further develop an intuition regarding the approach outlined in this paper. A more detailed practical experiment is given later in the paper.

The fact we can localize the pose of the object accurately (even up to an ambiguity determined by the number of Gaussians in the likelihood function) is quite novel given we only use class label measurements. However, we go further then this and show how class confusions (from certain view points) can even reduce the number of ambiguities.

Consider an object o_1 located at $\mathbf{x}_1 = [5\ 5]^\top$ with true orientation $\phi_1 = 45^o$. Consider two potential object classes c_1 and c_2 with defined likelihood functions

$$p(\widehat{c}_i, o_1|c_i, \phi_1, \mathbf{x}_1) = \frac{0.2495}{(2\pi)^{\frac{3}{2}}|\mathbf{\Sigma}_{1_i}|^{1/2}} \times \exp\left(\frac{-1}{2}\|\mathbf{\Sigma}_{1_i}^{\frac{-1}{2}}\left(\mathbf{p} - \mathbf{X}(\phi_1, \mathbf{x}_1)\mathbf{q}_{1_i}\right)\|_2^2\right) + \frac{0.2495}{(2\pi)^{\frac{3}{2}}|\mathbf{\Sigma}_{2_i}|^{1/2}} \exp\left(\frac{-1}{2}\|\mathbf{\Sigma}_{2_i}^{\frac{-1}{2}}\left(\mathbf{p} - \mathbf{X}(\phi_1, \mathbf{x}_1)\mathbf{q}_{2_i}\right)\|_2^2\right) \quad (25)$$

for both $i = 1$ and $i = 2$ (with $\mathrm{P}(\widehat{c}_i|c_i) = 1/2 - 0.001$ as a consequence). The mean parameters are given by $\mathbf{q}_{1_1} = \mathbf{q}_{1_2} = [0\ 10\ 0]^\top$ and $\mathbf{q}_{2_1} = \mathbf{q}_{2_2} = [0\ \ -10\ 0]^\top$. The false positive likelihoods are given by

$$p(\widehat{c}_i, o_1|c_j, \phi_1, \mathbf{x}_1) = \frac{0.2495}{(2\pi)^{\frac{3}{2}}|\mathbf{\Sigma}_{1_{i,j}}|^{1/2}} \times \exp\left(\frac{-1}{2}\|\mathbf{\Sigma}_{1_{i,j}}^{\frac{-1}{2}}\left(\mathbf{p} - \mathbf{X}(\phi_1, \mathbf{x}_1)\mathbf{q}_{1_{i,j}}\right)\|_2^2\right) \quad (26)$$

with $i \neq j \in \{1, 2\}$ and $\mathbf{q}_{1_{i,j}} = [0\ 10\ 0]^\top$ (and $\mathrm{P}(\widehat{c}_j|c_i) = 1/2 - 0.001$). The variance is given by $\mathbf{\Sigma}_{i_j} = \mathbf{\Sigma}_{1_{i,j}} = \mathrm{diag}(10, 10, \pi/4)$ for all combinations of i and j. Now consider the class c_0 with

$$p(\widehat{c}_i, o_1|c_0, \phi_1, \mathbf{x}_1) = \mathrm{P}(\widehat{c}_i|c_0) = 0.001 \quad (27)$$

for $i \in \{1, 2\}$. The recognition system can return class measurements $\widehat{c}_1$ and $\widehat{c}_2$.

We plot $p(\widehat{c}_i, o_1|c_i, \phi_1, \mathbf{x}_1)$ and $p(\widehat{c}_i, o_1|c_j, \phi_1, \mathbf{x}_1)$ with $i \neq j \in \{1, 2\}$ and with $\vartheta_i = 0$ over $\mathbf{s} \in \mathbb{R}^2$ in Figure 3.

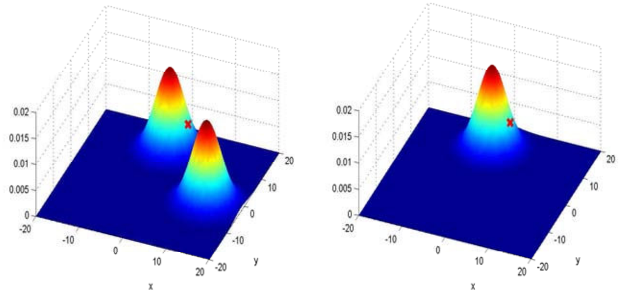


Fig. 3. The likelihoods $p(\widehat{c}_i, o_1|c_i, \phi_1, \mathbf{x}_1)$ and $p(\widehat{c}_i, o_1|c_j, \phi_1, \mathbf{x}_1)$ with $i \neq j \in \{1, 2\}$ evaluated at $\mathbf{x}_1 = [5\ 5]^\top$ and $\phi_1 = 45^o$. This shows the relationship between the robot position and the likelihoods.

In this example we assume $\mathbf{x}_1$ is known but the true object class and orientation ϕ_i is unknown. This is a reasonable approximation in many active object search problems[6]. In the next section, we consider a grid-based object search and orientation estimation problem where this assumption is explicitly realized.

The initial priors are thus $p(c_i, \phi_1, \mathbf{x}_1|o_1) = 1/(2\pi)$. We simulate measurements at a number of positions in space in order to examine their affect on the posterior densities.

Time 1

The robot position is given by $\mathbf{p} = \mathbf{X}(\phi_1, \mathbf{x}_1) \circ [2\ -10\ 0]^\top$. The measurements are given by $\mathbf{y}_1(1) = [\widehat{c}_1]^\top$. The posterior density functions are shown in Figure 4.

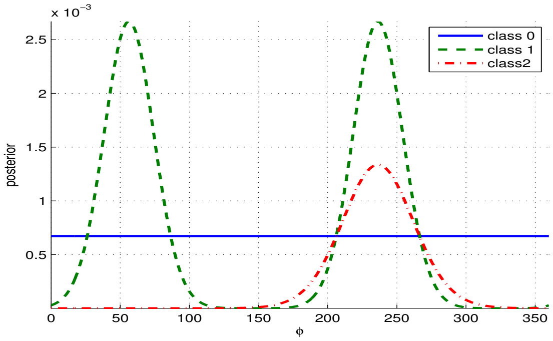


Fig. 4. The posteriors densities after the first measurement.

We know $\mathrm{P}(c_i|\mathbf{y}_j, o_j) = \int_{\mathbb{SO}(2)} p(c_i, \phi_j, \mathbf{x}_j|\mathbf{y}_j, o_j)\ d\phi_j$ and we can compute $\mathrm{P}(c_0|\mathbf{y}_1, o_1) = 0.4231$, $\mathrm{P}(c_1|\mathbf{y}_1, o_1) = 0.4244$ and $\mathrm{P}(c_2|\mathbf{y}_1, o_1) = 0.1524$ all at time 1. Then for the maximum class posterior estimate $\widetilde{c}_1$ we can compute the maximum $\mathrm{argmax}_{\phi_1}\, p(c_1, \phi_1, \mathbf{x}_1|\mathbf{y}_1(1), o_1)$ which is clearly (up to numerical tolerance) ambiguous with $\widetilde{\phi}_1 \approx 55^o$ and $\widetilde{\phi}_1 \approx 235^o$. The estimate of $\widetilde{c}_1$ is not overwhelmingly probable and the orientation estimate $\widetilde{\phi}_1$ is not exceedingly accurate since we have only employed a single measurement.

Time 2

The robot is at $\mathbf{p} = \mathbf{X}(\phi_1, \mathbf{x}_1) \circ [-2\ \ -10\ \ 0]^\top$. The measurements are given by $\mathbf{y}_1(2) = [\widehat{c}_1]^\top$. The posterior density functions are shown in Figure 5.

We compute $\mathrm{P}(c_0|\mathbf{y}_1, o_1) = 0.3019$, $\mathrm{P}(c_1|\mathbf{y}_1, o_1) = 0.5735$ and $\mathrm{P}(c_2|\mathbf{y}_1, o_1) = 0.1247$ at time 2. Then for the maximum class posterior estimate $\widetilde{c}_1$ we compute the maximum $\mathrm{argmax}_{\phi_1}\, p(c_1, \phi_1, \mathbf{x}_1|\mathbf{y}_1(1), o_1)$ which is again (up to numerical tolerance) ambiguous with $\widetilde{\phi}_1 \approx 45^o$ and $\widetilde{\phi}_1 \approx 225^o$. However, now the orientation estimate is accurate up to the ambiguity. The increased accuracy in the orientation (neglecting the ambiguity) is a result of the spatial averaging that occurs when observing the object from different robot positions (and this accuracy is quite interesting given we only physically measure the class label).

[6] For example, laser or stereo vision can be used to position objects in space in some scenarios but does not necessarily aid in the estimation of object class or orientation. In any case, we make this assumption here for simplicity and to make the example intuitively clear.

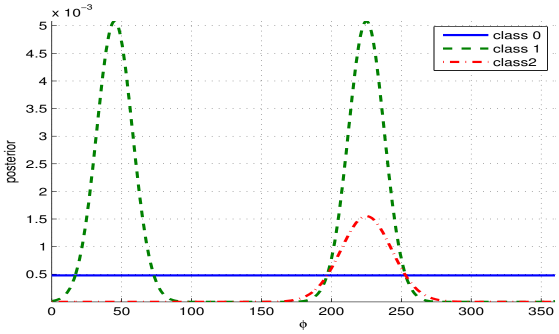


Fig. 5. The posteriors densities after the second measurement.

In the next time step we move to the other side of the object and show how confusion aids in removing the ambiguity.

$\boxed{\texttt{Time 3}}$

The robot position is given by $\mathbf{p} = \mathbf{X}(\phi_1, \mathbf{x}_1) \circ [0\ 10\ 0]^\top$. The measurements are given by $\mathbf{y}_1(3) = [\widehat{c}_1\ \ \widehat{c}_2]^\top$. The posterior density functions are shown in Figure 6.

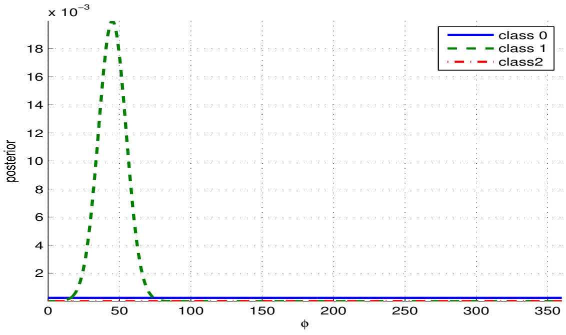


Fig. 6. The posteriors densities after the third measurement.

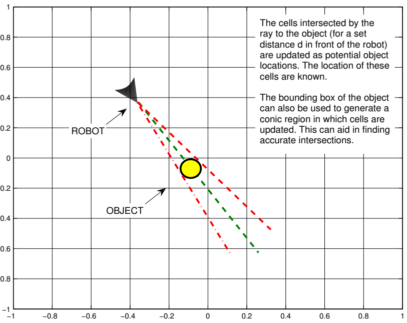
We compute $\mathrm{P}(c_0|\mathbf{y}_1, o_1) = 0.1502$, $\mathrm{P}(c_1|\mathbf{y}_1, o_1) = 0.8498$ and $\mathrm{P}(c_2|\mathbf{y}_1, o_1) = 1.8 \times 10^{-5}$ at time 3. Then for the maximum class posterior estimate $\widetilde{c}_1$ we compute the maximum $\mathrm{argmax}_{\phi_1}\, p(c_1, \phi_1, \mathbf{x}_1|\mathbf{y}_1(1), o_1)$ which is now unique $\widetilde{\phi}_1 \approx 45^o$. The orientation estimate is non-ambiguous in this case since we exploited inter-class confusion.

Note that we have estimated the orientation quite accurately using only measurements of the object class label and a pre-defined heuristic spatial likelihood function. We believe this is a novel result in the sense of minimalistic sensing[7].

IV. Grid-Based Object Classification and Orientation Estimation

Consider a grid on $\mathbb{R}^2$ denoted by $\mathcal{G}$. For simplicity, we assume the grid $\mathcal{G}$ consists of n_g grid squares of uniform size (the generalization to nonuniform grid cells is straightforward). Each grid square is denoted by $g_i \in \mathcal{G}$ and can be characterized by the center point $\mathbf{g}_i \in \mathbb{R}^2$. We are interested in assigning to each cell g_i the probability density $p(c_i, \phi_j, \mathbf{x}_j|\mathbf{y}_j, g_j)$ from which we can determine the probability $\mathrm{P}(c_i|\mathbf{y}_j, g_j)$ via marginalization. In fact, for each cell we assign n_c such probability densities - one for each class. Then $\sum_i \mathrm{P}(c_i|\mathbf{y}_j, g_j) = 1$ where $i = \{0, \ldots, n_c\}$ for each cell. In practice a lot of the cells will be dominated by the probability value $\mathrm{P}(c_0|\mathbf{y}_j, g_j)$.

In this scenario, $\mathbf{x}_j$ is the location of the j'th cell g_j and is known. If we imagine a robot located at $\mathbf{s}$ with ϑ_j the direction to o_j defines a ray which we limit to the length d. We update the set of cells $\{g_j\}$ that intersect such a ray using the posterior density formula given in the previous sections. The value $\mathbf{x}_j$ is taken as the cell center $\mathbf{g}_j \in \mathbb{R}^2$ and thus cells close or far from the robot (along the ray) are likely to be estimated as c_0. We could also define a conic region, e.g. by defining two rays using the bounding box of the object in the image, and then update the cells which intersect the conic region, e.g. see Figure 7.



The cells intersected by the ray to the object (for a set distance d in front of the robot) are updated as potential object locations. The location of these cells are known.

The bounding box of the object can also be used to generate a conic region in which cells are updated. This can aid in finding accurate intersections.

Fig. 7. An example grid environment.

The grid-based estimation problem follows closely the examples given in the last subsection where the location of each cell is known and is analogous to the location of an object position. For simplicity we have assumed cell independence. It is possible to relax this assumption but there are difficulties in doing so that are beyond the scope of this paper. For the grid-based scenario we will examine a practical experiment which is outlined in the next section.

V. Experimental Results

The robot considered in the experiments is equipped with a Point Grey Flea stereo camera (only one camera used in the experiment) on top of a Pioneer P3X robot base; see Figure 8.

We use FERNS as an object class detector [11], [12]. The robot position is computed from only odometry and the grid organization is known (each grid cell is 2 square decimeters).

The robot is in a room with three objects, o_1 is a box containing physics books and o_2 and o_3 are identical boxes containing robot parts. The boxes are located as shown in Figure 8. All objects have the same CAS lab logo on one of their sides and cannot be differentiated based on the class returns when viewed from this side. We call this *the confusion side* of the object. On the polar opposite side, o_1 exhibits a label indicating the box contains physics books whereas both o_2 and o_3 contain identical labels indicating

[7] The sequence of measurements (and confusions) affect the evolution of the posterior densities in interesting ways but we cannot explore all the cases here. In the experimental section more examples are given.

they contain robot parts. Views of this distinguishing box label are said to be of the *non-confusion side* of the object. The true orientations for o_1, o_2 and o_3 are $\phi_1 = 255°$, $\phi_2 = 315°$ and $\phi_3 = 180°$.

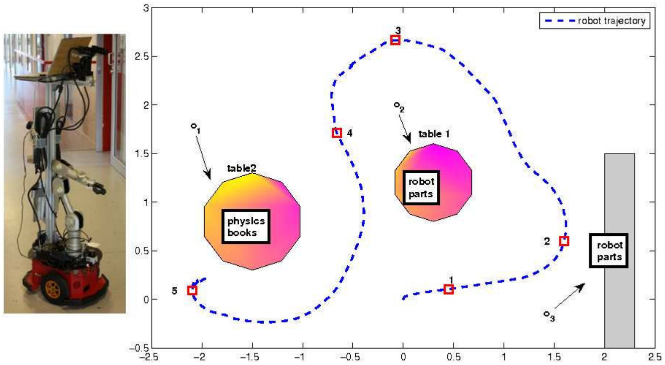


Fig. 8. [Left] CogX robotic platform and [Right] the robot trajectory and layout of the environment.

The robot starts at $(0,0)$ and follows the trajectory shown in Figure 8. Numbered positions in Figure 8 are where the robot takes a class label measurement. The non-object class, physics books box and robot parts box are labeled as c_0, c_1 and c_2 respectively. In this scenario the units are decimeters. As such, the likelihood functions $p(\widehat{c}_i, o_j | c_i, \phi_j, \mathbf{x}_j)$ for $i \in \{1,2\}$ and $j \in \{1,2,3\}$ are identical to those defined in (25) in the simulated example problem. Similarly, $p(\widehat{c}_i, o_j | c_k, \phi_j, \mathbf{x}_j)$ for $i \neq k \in \{1,2\}$ are identical to the likelihood functions defined in (26). Finally, $p(\widehat{c}_i, o_1 | c_0, \phi_1, \mathbf{x}_1)$ for $i \in \{1,2\}$ is identical to the function defined in (27). The recognition system can of course return class measurements $\widehat{c}_1$ and $\widehat{c}_2$.

A. Orientation Estimation at the Correct Grid Cell

The estimation algorithm in this section is run over a grid as discussed in the last section. However, to visualize the orientation estimate's density we need to essentially look at an individual cell. Thus, in this subsection we examine the orientation estimate in the practical experiment at the true object grid cell. Later we examine the grid map for the environment and show the distribution of the class label probabilities over a number of cells.

At point 1, the robot detects o_2 on its confusion side, i.e. both $\widehat{c}_1$ and $\widehat{c}_2$ are measured. The resulting orientation estimates for each class are shown in Figure 10 part (a). Since both $\widehat{c}_1$ and $\widehat{c}_2$ are detected and no further information is available, the probability estimates for both classes are equal but the maximum a posterior orientation estimate is non-ambiguous. The orientation estimate $\widetilde{\phi}_2 \approx 317°$ which is relatively close to the true orientation estimate.

At point 2, the robot detects o_3 from a non-confusion side; see Figure 9. The class measurement is only $\widehat{c}_2$ since the observed side is a discriminative one. However notice that the orientation estimate is multi-modal with $\widetilde{\phi}_3 \approx 3°$ and $\widetilde{\phi}_3 \approx 183°$. Since this box is on a shelf against a wall, it is not possible to observe it from other sides. We will show in the following how observing the confusion side improves the overall orientation estimate.

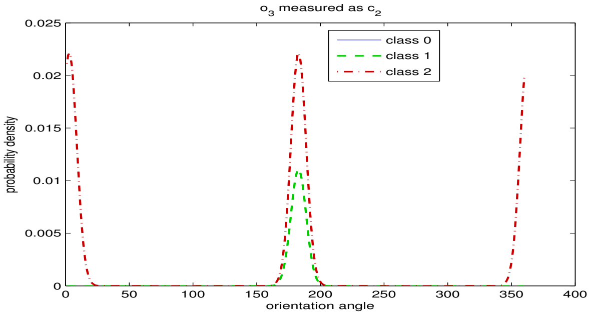


Fig. 9. The robot measures c_2 only. Note that the distribution is multi-modal. No further measurements are taken of this object.

At point 3, the robot observes a non-confusion side of o_2, i.e. only $\widehat{c}_2$ is measured which is the true class of o_2; see Figure 10 part (b). Notice that the probability over ϕ_2 for the class c_1 has dropped and will continue to do so as more measurements of $\widehat{c}_2$ are acquired.

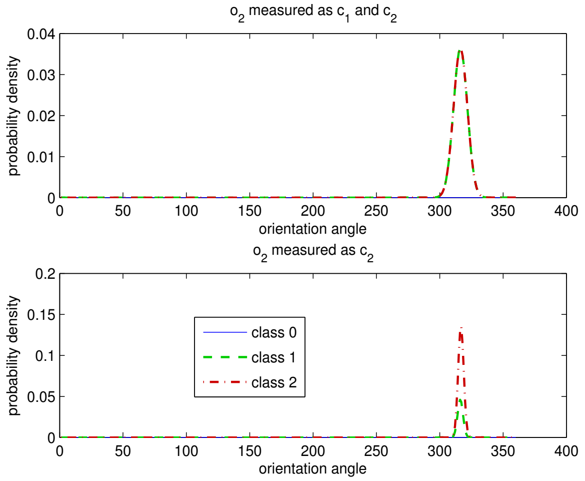


Fig. 10. [Top] The object is first seen from its confusion side. This is not enough to determine the class therefore class estimates are equal. [Bottom] The observation from its non-confusion side helps improving the class estimates.

At point 4, the robot detects the o_1 from its non-confusion side, i.e. only $\widehat{c}_1$ is measured which is the true class of o_1. As with the first measurement of o_3, we have two peaks for the detected class shown in Figure 11. The orientation estimate is given by $\widetilde{\phi}_1 \approx 57°$ and $\widetilde{\phi}_1 \approx 237°$. The maximum class estimate is $\widetilde{c}_1$.

At point 5, the robot observes o_1 from its confusion side. In this case since the object has been detected once from its non-confusion side, the probability of o_1 being of class c_1 is now much higher and the orientation estimate is now non-ambiguous with $\widetilde{\phi}_1 \approx 258°$ as shown in Figure 11. We now see that the confusion side helps to eliminate one of the peaks in the orientation estimate and the spatial likelihood function has helped the estimate converge to an accurate value.

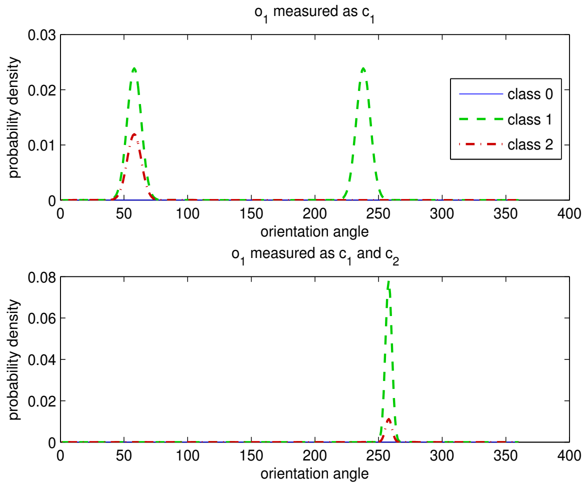


Fig. 11. [Top] The robot first measures c_1 and then [Bottom] both c_1 and c_2. Notice even though the confusion of the second measurement improves the orientation estimation.

B. Class Probability Estimation over the Grid Map

As described previously, the estimation algorithm is executed over a grid where the detected object rays define a set of cells updated after each measurement. Each cell is associated with an orientation and class density (for all possible classes). As an example, the marginalized probabilities for c_0, c_1 and c_2 are visualized in Figure 12 for class measurements of o_2 (recall we have assumed that particular class measurements can be assigned to the correct object).

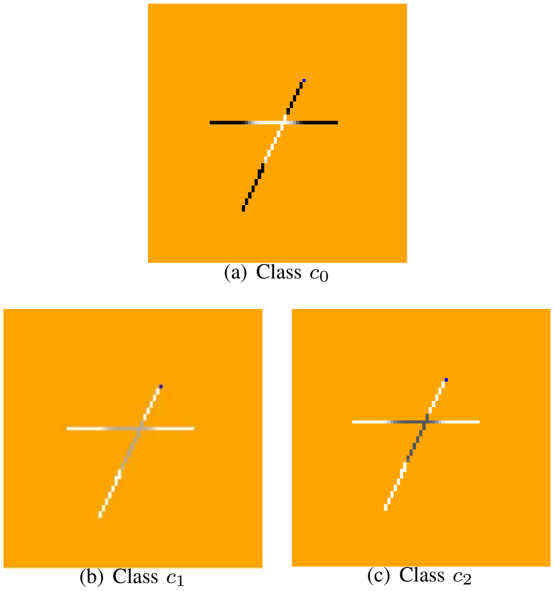
In this particular snapshot, o_2, which is of true class c_2, is seen from two positions (points 3 and 4 in Figure 8) and two rays are cast. The gray shading in each picture along the rays represents the probability $\mathrm{P}(c_i, o_2|y_2)$ for each respective class. We have normalized the shading so $\mathrm{P}(c_i, o_2|y_2) = 0$ is pure white while $\mathrm{P}(c_i, o_2|y_2) = 1$ is pure black. The orange background is used here to simplify visualization but can be thought of as the initial prior class probabilities for all cells (i.e. equal priors for all classes) and remains valid since these cells are not updated given only these measurements.

In Figure 12 we can note the probability $\mathrm{P}(c_2, o_2|y_2)$ along the rays increases in magnitude up until the grid cells located at the approximate object location, i.e. the intersection point of the rays. It then decreases as expected. Similarly, $\mathrm{P}(c_o, o_2|y_2)$ decreases in magnitude along the rays until the intersection where it is almost zero and then begins to increase as expected further away from the object.

VI. Conclusion

We have provided a solution to the problem of simultaneous object class and pose estimation using a generic object classifier and a spatially dependent measurement likelihood model. Our novelty is the ability to estimate both the class and pose of the objects in the environment given only measurements of the object class label from a generic classifier.

We believe the heuristics behind the design of the likelihood functions are realistic. However, one practically and theoretically interesting direction for future work includes the development of reinforcement-like learning algorithms for estimating the likelihood function parameters online. Another interesting direction for future work involves the design of control algorithms for actively searching the environment in order to maximize the information gain.



(a) Class c_0

(b) Class c_1

(c) Class c_2

Fig. 12. An example distribution over a grid where o_2, which is of true class c_2, is seen from points 3 and 4 in Figure 8.

References

[1] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, August 1988.

[2] S. Vasudevan, S. Gchteraa, V. Nguyena, and R. Siegwart. Cognitive maps for mobile robots - An object based approach. *Robotics and Autonomous Systems*, 55(5):359–371, May 2007.

[3] P.J. Besl. Geometric modeling and computer vision. *Proceedings of the IEEE*, 76(8):936–958, August 1988.

[4] D.F. Dementhon and L.S. Davis. Model-based object pose in 25 lines of code. *Intl. Journal of Computer Vision*, 15(1-2):123–141, 1995.

[5] G. Dudek and C. Zhang. Vision-based robot localization without explicit object models. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA'96)*, pages 76–82, April 1996.

[6] M.A. Sipe and D. Casasent. Global feature space neural network for active computer vision. *Neural Computing and Applications*, 7(3):195–215, September 1998.

[7] C.-P. Lu, G.D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):610–622, June 2000.

[8] H. Murase and S.K. Nayar. Visual learning and recognition of 3-d objects from appearance. *Intl. Journal of Computer Vision*, 14(1):5–24, January 1995.

[9] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz. Appearance-based active object recognition. *Image and Vision Computing*, 18(9):715–727, June 2000.

[10] B.D.O. Anderson and J.B. Moore. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, N.J., 1979.

[11] M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'07)*, pages 1–8, June 2007.

[12] Mustafa Ozuysal, Michael Calonder, Vincent Lepetit, and Pascal Fua. Fast keypoint recognition using random ferns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(3):448–461, March 2009.