YORK UNIVERSITÉ
UNIVERSITY

redefine THE POSSIBLE.

A Theory of Active Object Localization

Alexander Andreopoulos

John K. Tsotsos

Technical Report CSE-2009-01

March 10 2009

Department of Computer Science and Engineering
4700 Keele Street Toronto, Ontario M3J 1P3 Canada

A Theory of Active Object Localization

Alexander Andreopoulos, John K. Tsotsos
Dept. of Computer Science and Engineering
Centre for Vision Research, York University
Toronto, Ontario, Canada
{alekos, tsotsos}@cse.yorku.ca

Abstract

We present some theoretical results related to the problem of actively searching for a target in a 3D environment, under the constraint of a maximum search time. We define the object localization problem as the maximization over the search region of the Lebesgue integral of the scene structure probabilities. We study variants of the problem as they relate to actively selecting a finite set of optimal viewpoints of the scene for detecting and localizing an object. We do a complexity-level analysis on the problems, by showing that in the best case scenario, the problems have high order pseudo-polynomial running times or are NP-Complete. We study the tradeoffs of localizing vs. detecting a target object, using single-view and multiple-view recognition, under imperfect dead-reckoning and an imperfect recognition algorithm. We use these results to propose a set of sufficient properties that efficient and reliable active object localization algorithms should satisfy.

1. Introduction

In one of the earliest known treatises on vision [1], Aristotle describes vision as a passive process that is mediated by what he refers to as the "transparent" ($\delta\iota\alpha\varphi\alpha\nu\acute{\varepsilon}\varsigma$), an invisible property that allows the sense organ to become like the actual form of the visible object. Much has been learned since then and today, a popular definition is that vision is the process of discovering from images what is present in the world and where it is [10]. Within this context, four levels of tasks in the vision problem are discernible [17]:

- *Detection*: is a particular item present in the stimulus?
- *Localization*: detection plus accurate location of item.
- *Recognition*: localization of the items present in the stimulus plus their accurate description through their association with linguistic labels.
- *Understanding*: recognition plus role of stimulus in the context of the scene.

The concept of *active perception* or *active vision* was first introduced by Bajcsy [2], as "a problem of intelligent control strategies applied to the data acquisition process". Active control of a vision based sensor offers a number of benefits [19]. It allows us to: (i) Bring into the sensor's field of view regions that are hidden due to occlusion and self-occlusion. (ii) Foveate and compensate for spatial non-uniformity of the sensor. (iii) Increase spatial resolution through sensor zoom and observer motion that brings the region of interest in the depth of field of the camera. (iv) Disambiguate degenerate views due to finite camera resolution, lighting changes and induced motion [5]. (v) Deal with incomplete information and complete a task.

An active vision system's benefits must outweigh the associated execution costs [19]. The associated costs in an active vision system include:(i) Deciding the actions to perform and their execution order. (ii) The time to execute the commands and bring the actuators to their desired state. (iii) Adapt the system to the new viewpoint, find the correspondences between the old and new viewpoint and deal with the inevitable ambiguities due to sensor noise.

A number of active object detection, localization and recognition algorithms have been proposed over the years [3, 4, 6, 8, 11, 12, 13, 15, 16, 20, 21]. A smaller number of papers have dealt with issues related to the complexity and reliability of such systems [5, 9, 18, 19, 21]. Limited work exists on the complexity of search tasks and the effect that imperfect recognition and imperfect dead-reckoning has on object localization. In this paper, we argue that the problem is likely intractable, by proving that the active object localization problem is NP-Hard and by showing that the problem remains difficult at best, even under certain simplifying variants of the main problem. We study the tradeoffs of localizing vs. detecting a target object under single-view and multiple-view recognition schemes and show that there are a number of bias/variance/entropy relationships and tradeoffs between the reliability of target localization and target detection, that depend on the quality of the recognition algorithm used and the magnitudes of the correspondence or dead-reckoning errors. We exemplify the relevance of these results in practical computer vision applications, as first-principles based motivators for a set of properties that active object localization algorithms should satisfy.

2. Problem Formulation

Assumption 1. *We assume that exactly one instance of the target object exists in the scene.*

Definition 1. (Search Space) *The search space consists of a 3D region whose coordinates are expressed with respect to an inertial coordinate frame.*

Definition 2. (Target Map) *The target map is a discretization of the inertial coordinate frame into non-overlapping 3D cells coinciding with the search space. Each cell is assigned the probability of containing the target centroid.*

We use a set of positive integers, $C \triangleq \{1, 2, ..., |C|\}$, to index each cell in the target map. Notice, that, since we assume a single target object exists in the scene, the target map cell values sum to one.

Definition 3. (Scene Sample Function) *A scene sample function $\mu_v(\vec{x})$ denotes the sensor output, where v represents the values assigned to the controllable sensor parameters (*e.g.*, coordinate frame, zoom, focus) and $\vec{x}$ is an index into the scene sample (*e.g.*, in the case of greyscale images $\vec{x} = (i, j)$ can denote a pixel index).*

We define a probability space $\Upsilon = (X_1, \Sigma_1, p_1)$ for the sensor parameter states, where $v \in X_1$ denotes a sensor parameter state, Σ_1 is a σ-algebra of X_1 and p_1 is a probability measure on X_1 whose support includes all states v that have a non-zero probability of occuring in the search space. Similarly, for each v, we define a probability space $\Upsilon(v) = (X_v, \Sigma_v, p_v)$ with $p_v(\mu_v(\vec{x})) > 0$ for each $\mu_v(\vec{x}) \in X_v$, denoting the probability of occurence of the corresponding scene sample function given sensor parameter values v. The underlying probability measure, models the sensed scene uncertainty (*e.g.,* image noise, varying illumination conditions, dead-reckoning errors, etc.) and it is largely unknown and difficult to model in practice. Since we do not know the distribution of p_1, p_v, we approximate them by using a finite sample of optimally selected v, μ_v.

Definition 4. (Sequence Cost) *Given a sequence $v_1, ..., v_n$ of sensor parameter states, the cost $T(n)$ associated with executing the sequence is given by $T(n) \triangleq T(n-1) + \mathbf{t_o}(v_1, ..., v_n)$, where $\mathbf{t_o}(v_1, ..., v_n) > 0$ denotes the cost of moving to state v_n given all previous states and $T(1)$ is the cost of reaching state v_1 from the initial sensor state.*

We define the *3D object localization* and *constrained active object localization* (CAOL) problems as follows:

Definition 5. (3D Object Localization) *Find the cell $\hat{i}_t = \arg\max_i \int p(c_i|\mu_v(\vec{x}))dp_v dp_1$, where we are taking the Lebesgue integrals[14] over Υ and $\Upsilon(v)$ and c_i denotes the event that the target object's centroid is in cell i. $p(c_i|\mu_v(\vec{x}))$ is a recognition algorithm depending on v, μ_v. If $p(c_i|\mu_v(\vec{x}))$ is a "good" algorithm, $\hat{i}_t = i_t$, where i_t is defined as the cell containing the target's centroid.*

Definition 6. (Constrained Active Object Localization) *Find the cell $\hat{i}_t \in C$ maximizing $p(c_{\hat{i}_t}|\mu_{v_n}(\vec{x}), ..., \mu_{v_1}(\vec{x}))$ across all $n > 0$, all sequences $v_1, ..., v_n$ of sensor states and all corresponding $\mu_{v_1}, ..., \mu_{v_n}$, under the constraint $T(n) \leq T'$, where T' is a search cost bound.*

Solutions to the CAOL problem must compensate for (i) our limited knowledge on Υ, $\Upsilon(v)$ and (ii) the need to minimize sensor movements, by finding a finite sample $\mu_{v_n}(\vec{x}), ..., \mu_{v_1}(\vec{x})$ that best samples the unknown probability spaces without exceeding the maximum alloted search cost. Even if we know the distributions of the probability spaces Υ, $\Upsilon(v)$, eliminating point (i) and potentially even making $p(c_i|\mu_{v_n}(\vec{x}), ..., \mu_{v_1}(\vec{x}))$ a function of v_1,...,v_n, the problem remains intractable. As we show later, the CAOL problem belongs to the class of NP-Hard problems [7], implying that there is no known polynomial time algorithm that solves the problem. One can attempt to make it tractable by using variants of the problem:

Definition 7. (Constrained Active Object Localization: Variant 1) *Find a sequence $v_1, ..., v_n$ of sensor states and the cells $\hat{i}_t \in C$ satisfying $p(c_{\hat{i}_t}|\mu_{v_n}(\vec{x}), ..., \mu_{v_1}(\vec{x})) \geq \theta$ and $T(n) \leq T'$ for some $\mu_{v_1}, ..., \mu_{v_n}$, where T' is a search cost bound and θ is a probability threshold.*

Definition 8. (Constrained Active Object Localization: Variant 2) *Find the cell $\hat{i}_t \in C$ maximizing $p(c_{\hat{i}_t}|\mu_{v_n}(\vec{x}), ..., \mu_{v_1}(\vec{x}))$ across all $n > 0$, all sequences $v_1, ..., v_n$ of sensor states and all corresponding $\mu_{v_1}, ..., \mu_{v_n}$, under the constraint $T(n) \leq T'$, where T' is a search cost bound and each movement cost $\mathbf{t_o}(v_1, ..., v_i)$ is bounded from below by a positive non-zero constant C'.*

Theorem 1. (Simplified Bayesian Updating) *Assume $p(\mu_{v_n}|c_i, \mu_{v_{n-1}}, ..., \mu_{v_1}) = p(\mu_{v_n}|c_i)$. Then, $p(c_i|\mu_{v_n}, ..., \mu_{v_1}) = \frac{p(c_i|\mu_{v_{n-1}},...,\mu_{v_1})p(\mu_{v_n}|c_i)}{\sum_j p(c_j,\mu_{v_n}|\mu_{v_{n-1}},...,\mu_{v_1})}$.*

Proof. $p(c_i|\mu_{v_n}, ..., \mu_{v_1})p(\mu_{v_n}, ..., \mu_{v_1})= p(c_i, \mu_{v_{n-1}}, ..., \mu_{v_1})p(\mu_{v_n}|c_i) \Leftrightarrow p(\mu_{v_n}|c_i, \mu_{v_{n-1}}, ..., \mu_{v_1})= p(\mu_{v_n}|c_i)$. Notice also that $\sum_j p(c_j, \mu_{v_n}|\mu_{v_{n-1}}, ..., \mu_{v_1}) = \sum_j p(\mu_{v_n}|c_j)p(c_j|\mu_{v_{n-1}}, ..., \mu_{v_1})$. □

When we are not using the simplifying assumption stated in Theorem 1, we say we are using *normal Bayesian updating*. Theorem 1 assumes that the scene sample functions are conditionally independent given the cell i where the target is centred. By Assumption 1, exactly one instance of the target exists in the scene, which implies that event c_i is sufficient to determine which regions of μ_{v_n} (if any) correspond to the projection of the target object on the image plane and which regions correspond to the background. We are implicitly assuming that $p(\mu_{v_n}|c_i)$ denotes a generative modeling of the recognition algorithm's resultant binary segmentation into the foreground (target position) and the background, based on a single view. Similarly $p(c_i|\mu_{v_n},...,\mu_{v_1})$ denotes the corresponding probability of event c_i, based on the bayesian fusion of multiple-views μ_{v_n},...,μ_{v_1}. Notice that for a uniform prior $p(c_i)$, $\arg\max_i p(\mu_{v_n}|c_i) = \arg\max_i p(c_i|\mu_{v_n})$. The greater the uncertainty implicit in spaces $\Upsilon(v)$, the weaker the assumption of conditional independence becomes, due to increased sources of error. Nevertheless, it is convenient to use Theorem 1 to model various localization and detection tradeoffs.

In the next section, we prove that if we know the distributions of Υ, $\Upsilon(v)$ and under normal Bayesian updating, Variant 1 of the CAOL problem (Def.7) and the corresponding detection problem are NP-Hard and NP-Complete respectively. It is easy to see that Def.7 is reducible to the similarly discretized version of Def.6 and thus, the CAOL problem is NP-Hard. Variant 2 of the problem, has a high-order pseudo-polynomial solution: Since there are at most $\lfloor \frac{T'}{C'} \rfloor$ sensor settings to execute within time T', an enumeration and evaluation of all candidate solutions, runs in $\Omega(m^{\lfloor \frac{T'}{C'} \rfloor})$, where m is the total number of possible states. But this solution remains exponential in terms of the size of T'. Using a reduction from Def.7, we notice that Def.8 is NP-Hard and if we add to Def.7 the minimum cost constraint of Def.8, the resulting problem remains NP-Hard—the reductions involve setting C' to the minimum sensor state pair cost. We could also approach the localization problem by thresholding the generative probability $p(\mu_{v_n}(\vec{x})|c_i)$ rather than the discriminative probability $p(c_i|\mu_{v_n}(\vec{x}),...,\mu_{v_1}(\vec{x}))$. Ye [21] uses a binary classifier with a presumed zero false positive rate, to show that a similar problem is NP-Complete.

3. The Constrained Active Object Localization Problem:Variant 1, is NP-Hard

To analyze the complexity of the constrained active object localization problem when we know the distributions of Υ, $\Upsilon(v)$, we first reformulate the problem into the corresponding detection problem, taking into account the finite precision of floating point arithmetic, and the finite set V that is necessary to represent the space of scene sample functions (X_v) achievable across the sensor parameter states (X_1). Let $\mathbb{Q}^+ \triangleq \{\frac{p}{q} : p, q \in \mathbb{Z}^+\}$ denote the set of positive rational numbers. We model each probability by a non-negative rational in $\mathbb{Q}_1^+ \triangleq \{x \in \mathbb{Q}^+ \cup \{0\} : x \leq 1\}$.

Definition 9. (Valid Sequence) *Let* v_i'=$(v_{\pi_i(1)},...,v_{\pi_i(l(i))})$ *denote an ordered set of length* $l(i)$, *where* π_i:$\mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ *is a one-to-one mapping. A sequence* $v_{i_1}',...,v_{i_n}'$ *of ordered sets is valid if* $l(i_1) = 1$, $l(i_{k+1}) = l(i_k) + 1$ *for each* $1 \leq k \leq n-1$ *and* $\pi_{i_k}(j) = \pi_{i_{k+1}}(j)$, $\forall j$ $1 \leq j \leq l(i_k)$.

We define an ordered set of length zero as $v_0' \triangleq ()$. For any ordered set $v_i' = (v_{\pi_i(1)},...,v_{\pi_i(l(i))})$ let $v_i'(v_c) \triangleq (v_{\pi_i(1)},...,v_{\pi_i(l(i))},v_c)$. Also $v_0'(v_c) \triangleq (v_c)$ and $i_0 \triangleq 0$.

Definition 10. (Π_{j_t} :Constrained Active Object Detection Problem : Variant 1)
INSTANCE: A finite set $V = \{v_1,...,v_{|V|}\}$. *A cost constraint* $B' \in \mathbb{Q}^+$, *and a cost function* $C(v_i') \in \mathbb{Q}^+$ *where* $v_i' = (v_{\pi_i(1)},...,v_{\pi_i(l(i))})$, $i \in \mathbb{Z}^+$, $v_{\pi_i(1)},...,v_{\pi_i(l(i))} \in V$. $S' \in \mathbb{Z}^+$ *denoting the number of cells in the target map. A function* $f_1(v_i', j) \in \mathbb{Q}_1^+$ *such that for any ordered set* v_i' *and any* $1 \leq j \leq S'$, $\sum_{v_c \in V} f_1(v_i'(v_c), j) = 1$. *A function* $f_2(v_i', j) \in \mathbb{Q}_1^+$ *defined for* $1 \leq j \leq S'$, *such that* $\sum_{j=1}^{S'} f_2(v_i', j) = 1$ *for all ordered sets* v_i' *and* $f_2(v_{i_n}', j) \triangleq f_2(v_0', j) \prod_{k=1}^{n} \frac{f_1(v_{i_k}', j)}{\sum_{c=1}^{S'} f_2(v_{i_{k-1}}', c) f_1(v_{i_k}', c)}$. *A recognition threshold* $\theta \in \mathbb{Q}_1^+$. *A query cell* $1 \leq j_t \leq S'$.
QUESTION: Is there a valid sequence $v_{i_1}',...,v_{i_n}'$ *so that* $\sum_{k=1}^{n} C(v_{i_k}') \leq B'$ *and* $f_2(v_{i_n}', j_t) \geq \theta$*?*

Definition 11. ($\overline{\Pi}$:Constrained Active Object Localization Problem : Variant 1)
INSTANCE: Same as in Π_{j_t} *(*j_t *can be arbitrary). We use a bar to differentiate the input variables from those of* Π_{j_t}.
TASK: Find a valid sequence $v_{i_1}',...,v_{i_n}'$ *and the corresponding cells* j, $1 \leq j \leq \bar{S}'$, *which satisfy* $\sum_{k=1}^{n} \bar{C}(v_{i_k}') \leq \bar{B}'$ *and* $\bar{f}_2(v_{i_n}', j) \geq \bar{\theta}$.

As $\theta, \bar{\theta}$ decrease, the expected running times of Π_{j_t}, $\overline{\Pi}$ do not increase (*e.g.*, for $\theta, \bar{\theta} = 0$, solutions in $O(|V|)$ are trivial to find). Notice that for $\bar{\theta} > \frac{1}{2}$, there is at most one cell that $\overline{\Pi}$ can output. We quote the Knapsack problem (an NP-Complete problem) as given by Garey and Johnson [7]:

Definition 12. (Π' :Knapsack Problem)
INSTANCE: A finite set U, *a "size"* $s(u) \in \mathbb{Z}^+$ *and a "value"* $w(u) \in \mathbb{Z}^+$ *for each* $u \in U$, *a size constraint* $B \in \mathbb{Z}^+$, *and a value goal* $K \in \mathbb{Z}^+$.
QUESTION: Is there a subset $U' \subseteq U$ *such that* $\sum_{u \in U'} s(u) \leq B$ *and* $\sum_{u \in U'} w(u) \geq K$*?*

Π_{j_t} is in NP, since any candidate solution is verifiable in polynomial time. We assume $\frac{K}{\sum_{u \in U} w(u)} \leq 1$ since otherwise $\nexists U' \subseteq U$ that satisfies Π'. We define a mapping f from Π' to Π_{j_t} for which Π' is true iff Π_{j_t} is true:

1. $V \leftarrow U$
2. $B' \leftarrow B$
3. $C(v_i') = s(v_{\pi_i(l(i))})$

4. $S' \leftarrow 2$
5. $\theta \leftarrow \frac{K}{\sum_{u \in V} w(u)}$
6. We need to define $f_1(v'_i, j)$ and $f_2(v'_i, j)$ for all ordered sets v'_i that are composed of elements in V and all j, $1 \leq j \leq S'$, such that f_1, f_2 satisfy their preconditions stated in Π_{j_t}.

For each distinct set $U' \subseteq V$ and each distinct ordering o of the elements in U', we assume $d(U', o) \in \mathbb{Z}^+$ is unique and denotes the identifier of the corresponding ordered set $v'_{d(U',o)} = (v_{\pi_{d(U',o)}(1)}, ..., v_{\pi_{d(U',o)}(l(d(U',o)))})$ where $l(d(U', o)) = |U'|$. Furthermore, $d(U', o, k)$, for $1 \leq k \leq l(d(U', o))$, denotes the ordered set composed of the first k elements of $v'_{d(U',o)}$ — *i.e.*, $v'_{d(U',o,k)} = (v_{\pi_{d(U',o)}(1)}, ..., v_{\pi_{d(U',o)}(k)})$ and $v'_{d(U',o,k)} = v'_j$ iff $d(U', o, k) = j$. For any ordering o and set $U' = \{v_{\pi_{d(U',o)}(1)}, ..., v_{\pi_{d(U',o)}(l(d(U',o)))}\} \subseteq V$, we need to define $f_1(v'_{d(U',o,k)}, j)$ and $f_2(v'_{d(U',o,k)}, j)$ for all $1 \leq k \leq l(d(U', o))$. We also need to make sure $f_1(v'_{d(U',o,k)}, j)$ and $f_2(v'_{d(U',o,k)}, j)$ satisfy the requirements set in the definition of Π_{j_t} and only depend on j and the first k parameters of $v'_{d(U',o)}$. For each instance of Π' we define f_2 in Π_{j_t} by

$$f_2(v'_i, j) = \begin{cases} \dfrac{\sum_{k=1}^{l(i)} w(v_{\pi_i(k)})}{\sum_{u \in V} w(u)} & \text{if } j = j_t \\ \dfrac{1}{S'-1}\left(1 - \dfrac{\sum_{k=1}^{l(i)} w(v_{\pi_i(k)})}{\sum_{u \in V} w(u)}\right) & \text{otherwise} \end{cases}$$

Since $\sum_{j=1}^{S'} f_2(v'_i, j) = 1$, $f_2(v'_i, j)$ satisfies the requirements in Π_{j_t}. Notice from Def.10 that if $f_2(v'_{i_k}, j) = 1$, then $f_1(v'_{i_k}, j) \neq 0$. Also, if $0 < f_2(v'_{i_k}, j) < 1$, then $0 < f_2(v'_{i_{k-1}}, j) < 1$. From the definition of Π_{j_t}, for each subset U', each ordering o and each $1 \leq k \leq l(d(U', o))$, we want to define f_1 so that

$$f_2(v'_{i_k}, j) = \frac{f_2(v'_{i_{k-1}}, j) f_1(v'_{i_k}, j)}{\sum_{j'=1}^{S'} f_2(v'_{i_{k-1}}, j') f_1(v'_{i_k}, j')} \tag{2}$$

where $i_k = d(U', o, k)$, $1 \leq k \leq l(d(U', o))$, is used to denote a valid sequence of ordered sets. From Lemma 1 below, we know that for each sensor setting v'_{i_k} and $\forall j$, there exists an assignment to function $f_1(v'_{i_k}, j)$ that satisfies Eq.(2) and depends only on the parameters v'_{i_k}, j — *i.e.*, given parameters v'_{i_k} and j, f_1 is independent of set U'. Also Eq.(2) is independent of scaling factors applied on f_1, implying that we can assume that $\sum_{v_c \in V} f_1(v'_i(v_c), j) = 1$ as wanted. We see that mapping f runs in polynomial time.

We now show that there exists a valid sequence v'_{i_1},...,v'_{i_n} that satisfies Π_{j_t}, iff $\exists U' \subseteq U$ that satisfies Π': If Π_{j_t} holds, $f_2(v'_{i_n}, j_t) \geq \theta \Rightarrow \sum_{u \in U'} w(u) \geq K$ where $U' = \{v_{\pi_{i_n}(1)}, ..., v_{\pi_{i_n}(l(i_n))}\} \subseteq U$. Conversely, assume that for a subset $U' \subseteq U$ problem Π' holds. Choose an arbitrary ordering o and let $i_k = d(U', o, k)$, $1 \leq k \leq l(d(U', o)) = n$. We see that $f_2(v'_{i_n}, j_t) \geq \theta$. The converse direction of the proof holds regardless of the ordering assigned to U'. Regardless of the ordering o assigned to U', $\sum_{k=1}^{l(i_n)} C(v'_{i_k}) \leq B'$ iff $\sum_{u \in U'} s(u) \leq B$, which proves that there is a subset U' satisfying Π' iff an ordered set satisfies Π_{j_t}. This proves that Π_{j_t}, under normal Bayesian updating, is NP-Complete.

To prove that $\overline{\Pi}$ is NP-Hard, we define a mapping from Π_{j_t} to $\overline{\Pi}$ as follows: $\bar{V} \leftarrow V$, $\bar{B}' \leftarrow B'$, $\bar{C}(v'_i) = C(v'_i)$, $\bar{S}' \leftarrow 2$, $\bar{\theta} \leftarrow \frac{2}{3}$, $\bar{f}_2(v'_{i_k}, 1) = \frac{2}{3}\mathrm{I}_{A(k)} + \frac{1}{2}\mathrm{I}_{\bar{A}(k)}$, $\bar{f}_2(v'_{i_k}, 2) = \frac{1}{3}\mathrm{I}_{A(k)} + \frac{1}{2}\mathrm{I}_{\bar{A}(k)}$, where $\mathrm{I}_X \in \{0, 1\}$ is an indicator function that takes a value of 1 iff boolean variable X is true and $A(k), \bar{A}(k)$ are true iff $f_2(v'_{i_k}, j_t) \geq \theta$ or $f_2(v'_{i_k}, j_t) < \theta$ respectively. By Lemma 1, this also implicitly defines $\bar{f}_1$. We see that Π_{j_t} holds iff $\overline{\Pi}$ finds a valid sequence that is satisfied by cell $j = 1$. This shows that $\overline{\Pi}$ is NP-Hard.

In the reduction from Π' to Π_{j_t}, each call to f_2 is in $O(|V|)$ and takes $O(|V| \cdot S')$ space to encode. We are making the implicit assumption that f_1, f_2 in Π_{j_t} and $\bar{f}_1$, $\bar{f}_2$ in $\overline{\Pi}$ have running times and encoding sizes that are polynomial functions of $|V|$, S' and $|\bar{V}|$, $\bar{S}'$ respectively, implying that the scene structure must exhibit a minimum degree of "non-randomness". From the above proofs and Lemma 1, we notice that $f_1(v'_{i_k}, j)$ and $\bar{f}_1(v'_{i_k}, j)$ correspond to $p(\mu_{v_k} | c_j, \mu_{v_{k-1}}, ..., \mu_{v_1})$. Only if $\bar{f}_1(v'_{i_k}, j)$ depended exclusively on j and $v_{\pi_{i_k}(l(i_k))}$, would this constitute a proof that Def.11 is NP-Hard under simplified Bayesian updating. $f_2(v'_0, j)$ and $\bar{f}_2(v'_0, j)$ denote the prior distributions of the target maps and are typically set to a uniform distribution.

Lemma 1. *Let* $\beta, \alpha_1, ..., \alpha_m \in \mathbb{Q}_1^+$ *such that* $\sum_{i=1}^m \alpha_i = 1$, *if* $\beta = 1$, *then* $\alpha_1 \neq 0$ *and if* $0 < \beta < 1$, *then* $0 < \alpha_1 < 1$. *If* $m > 1$, $\exists x_1, ..., x_m \in \mathbb{Q}_1^+$ *such that* $\frac{\alpha_1 x_1}{\sum_{i=1}^m \alpha_i x_i} = \beta$.

Proof. If $\beta = 1$, let $x_1 = 1$ and let $x_i = 0$ for $i \neq 1$. If $\beta = 0$, let $x_2 = 1$ and let $x_i = 0$ for $i \neq 2$. Otherwise, if $0 < \beta < 1$, assume $x_1 > 0$ and notice that $\frac{\alpha_1 x_1}{\sum_{i=1}^m \alpha_i x_i} = \beta \Leftrightarrow \alpha_1 - \beta\alpha_1 = \sum_{i=2}^m (\beta\alpha_i) y_i$, a linear equation of $y_i = \frac{x_i}{x_1}$. Since $0 < \beta < 1$, $0 < \alpha_1 < 1$ and consequently $\sum_{i=2}^m \alpha_i > 0$, which implies $\alpha_1 - \beta\alpha_1 > 0$ and $\sum_{i=1}^m \beta\alpha_i > 0$. Therefore, there exist $y_2, ..., y_n \geq 0$ which satisfy the linear equation. We leave it as an exercise for the reader to verify that for any $y_2, ..., y_n \in \mathbb{Q}^+ \cup \{0\}$, $\exists x_1, x_2, ..., x_n \in \mathbb{Q}_1^+$ $(x_1 \neq 0)$ which satify $y_i = \frac{x_i}{x_1}$. □

4. Localization vs. Detection

We formalize some of the tradeoffs of single-view and multiple-view recognition schemes for localizing and detecting a target object under simplified Bayesian updating and under a number of different sources of errors. In Sec. 4.1 we define and discuss the problems and in Sec. 4.2-4.3 we prove the respective theorems.

4.1. Definitions and Discussion

Definition 13. (Correspondence Error) *Any error in the calculation of the correspondence(s) between the index value $\vec{x}$ of a scene sample function $\mu_v(\vec{x})$ and the target map cell indices whose structure projects on $\vec{x}$.*

Definition 14. (Dead-Reckoning Errors) *We are dealing with dead-reckoning errors when there exists a rigid transformation $RT(\cdot)$ of the sensor's estimated coordinate frame with respect to the inertial coordinate frame of the search space, that corrects all correspondence errors without introducing any new correspondence errors.*

Definition 15. (Visibility) *Cell i is visible for state v_n, if it falls in the sensor's field of view and satisfies a set of necessary conditions for localizing a target centered in i, that only depend on the coordinates of a point in i and the depth map of μ_{v_n} with respect to the sensor coordinate frame.*

Definition 16. (Good Single-View Recognition) *We have good single-view recognition at step n if $p(\mu_{v_n}|c_{i_t})$ is not affected by changes to the inertial coordinate frame. Also, under dead-reckoning errors, $p(\mu_{v_n}|c_i) \geq p(\mu_{v_n}|\neg c_i)$ for all target map distributions at step $n-1$ iff $i \in \hat{V}(v_n)$ and $RT(i_t) = i$, or, $i \notin \hat{V}(v_n)$ and $RT(i_t) \notin \hat{V}(v_n)$.*

$RT(i_t)$ denotes the cell containing the transformation of the target's centroid under $RT(\cdot)$ (Def.14). $p(\mu_{v_n}|\neg c_i)$ is defined in Sec. 4.2. $V(v_n)$ is the ground truth of visible cells for μ_{v_n}, v_n and no correspondence errors, while $\hat{V}(v_n)$ denotes the calculated visible cells based on our estimate of the sensor coordinate frame and under no guaranty of perfect correspondences. Under perfect correspondences $\hat{V}(v_n) = V(v_n)$, but the converse does not hold. For good single-view recognition, as the correspondence errors increase, it is more likely that $p(\mu_{v_n}|c_{i_t}) < p(\mu_{v_n}|\neg c_{i_t})$. Def.16 implies that if $i_1, i_2 \notin \hat{V}(v_n)$, $i_3 \in \hat{V}(v_n)$ and $RT(i_t) \notin \hat{V}(v_n)$, $p(\mu_{v_n}|c_{i_1}) = p(\mu_{v_n}|c_{i_2})$ and $p(\mu_{v_n}|c_{i_3}) < p(\mu_{v_n}|c_{i_1})$. Also, if $RT(i_t) \in \hat{V}(v_n)$, $p(\mu_{v_n}|c_{RT(i_t)}) > p(\mu_{v_n}|c_j)\ \forall j \neq RT(i_t)$ (see Sec. 4.2).

Theorem 2. (Detection Tradeoff)
Assume $i_t \in V(v_1)$,...,$i_t \in V(v_n)$. Assume a uniform target map prior and good single-view recognition. Let $X_i^{(n)}$, $Y_i^{(n)}$ denote Bernoulli random variables with probability of success $p(c_i|\mu_{v_n}, ..., \mu_{v_1})$, $p(c_i|\mu_{v_n})$ respectively. Detection at step n is based on $\max_{j\in\hat{V}(v_n)} E(X_j^{(n)})$ or $\max_{j\in\hat{V}(v_n)} E(Y_j^{(n)})$ being above a given threshold.
(i) Given v_n, μ_{v_n}, single-view detection at step n is independent of dead-reckoning errors.
(ii) If $p(\mu_{v_n}|c_i) \leq p(\mu_{v_n}|\neg c_i)$, $E(X_i^{(n)}) \leq E(X_i^{(n-1)})$.
(iii) If $p(\mu_{v_n}|c_i) \geq p(\mu_{v_n}|\neg c_i)$, $E(X_i^{(n)}) \geq E(X_i^{(n-1)})$.
(iv) If $\hat{i}_t = \hat{j}_t$, $\hat{i}_t = \arg\max_{j\in C} E(Y_j^{(n)})$ and $\hat{j}_t = \arg\max_{j\in C} E(Y_j^{(n-1)})$, $E(X_{\hat{i}_t}^{(n)}) \geq E(X_{\hat{j}_t}^{(n-1)})$.
(v) If $\hat{i}_t \neq \hat{j}_t$, $\hat{i}_t = \arg\max_{j\in C} E(Y_j^{(n)})$ and $\hat{j}_t = \arg\max_{j\in C} E(Y_j^{(n-1)})$, then it is not necessarily the case that $E(X_{\hat{i}_t}^{(n)}) \geq E(X_{\hat{j}_t}^{(n-1)})$.

Case (iv) shows that with good correspondences, detection based on fusing multiple views becomes more reliable than single-view detection (since $\hat{i}_t, \hat{j}_t \in \hat{V}(v_n)$). Case (v) shows that under dead-reckoning errors, there is an increased likelihood that fusing multiple-views will lead to more false negative detections (since $\hat{i}_t, \hat{j}_t \in \hat{V}(v_n)$), and thus, single-view detection (case (i)) might be preferable when dead-reckoning errors occur. Despite the strong assumption of Def.16, correspondence or dead-reckoning errors make the detection problem significantly harder.

Definition 17. (Dual Support) *Let $x_i^{(n)} \triangleq p(c_i|\mu_{v_n}, ..., \mu_{v_1})$. A single-view recognition algorithm has dual support at step n if $\forall i$, $x_i^{(n)} \notin [\frac{1}{e}, \frac{1}{2}]$. Equivalently $\forall i$, $\frac{p(\mu_{v_n}|\neg c_i)}{p(\mu_{v_n}|c_i)} > \frac{x_i^{(n-1)}}{1-x_i^{(n-1)}}(e-1)$ or $\frac{p(\mu_{v_n}|c_i)}{p(\mu_{v_n}|\neg c_i)} > \frac{1-x_i^{(n-1)}}{x_i^{(n-1)}}$.*

Definition 18. (Flipped Cells) *We say that there exist flipped cells at step n, if there exist two cells i_1, i_2, such that $x_{i_1}^{(n-1)} > \frac{1}{2}$, $x_{i_2}^{(n-1)} < \frac{1}{2}$, $x_{i_1}^{(n)} = x_{i_1}^{(n-1)} - x_1 < \frac{1}{2}$, $x_{i_2}^{(n)} = x_{i_2}^{(n-1)} + x_2 > \frac{1}{2}$ for positive x_1, x_2.*

Under Def.16 and a uniform target map prior, flipped cells can only occur due to correspondence errors.

Definition 19. (Boundary Constraints) *We say that the cells in a set S satisfy the boundary constraints at step n if for each $i \in S$, $p(\mu_{v_n}|c_i) < p(\mu_{v_n}|\neg c_i)$ and*

$$p(c_i|\mu_{v_{n-1}}, ..., \mu_{v_1}) < \frac{p(\mu_{v_n}|\neg c_i) - \sqrt{p(\mu_{v_n}|c_i)p(\mu_{v_n}|\neg c_i)}}{p(\mu_{v_n}|\neg c_i) - p(\mu_{v_n}|c_i)},$$

or, $p(\mu_{v_n}|c_i) > p(\mu_{v_n}|\neg c_i)$ and

$$p(c_i|\mu_{v_{n-1}}, ..., \mu_{v_1}) > \frac{p(\mu_{v_n}|\neg c_i) - \sqrt{p(\mu_{v_n}|c_i)p(\mu_{v_n}|\neg c_i)}}{p(\mu_{v_n}|\neg c_i) - p(\mu_{v_n}|c_i)}.$$

Theorem 3. (Localization Tradeoff)
Assume C satisfies the boundary constraints at step n. Also assume a uniform prior distribution for the target map. Define $d_i^{(n)} \triangleq x_i^{(n-1)} - x_i^{(n)}$ and $r_{i,k}^{(n)} \triangleq \frac{d_i^{(n)}}{\sum_{j\neq k} d_j^{(n)}}$.
(i) Assume there are no flipped cells at step n and $\forall i$ $x_i^{(n-1)} \leq \frac{1}{2}$. Then, there exists a cell i_1 for which $x_{i_1}^{(n)} > \frac{1}{2}$. Furthermore, if $x_{i_1}^{(n-1)} > \prod_{i\neq i_1}(x_i^{(n-1)})^{r_{i,i_1}^{(n)}}$, the target map entropy at step n is smaller than it is at step $n-1$.
(ii) If $x_{i_1}^{(n-1)} > \frac{1}{2}$ for some cell i_1, there exists a cell j_1, which does not have to equal i_1, such that $x_{j_1}^{(n)} > \frac{1}{2}$.
(iii) If there are no flipped cells at step n and there exists a

cell i_1 satisfying $x_{i_1}^{(n-1)} > \frac{1}{2}$ and $x_{i_1}^{(n)} > \frac{1}{2}$, then, the target map entropy at step n is smaller than it is at step $n-1$.
(iv) If there exist flipped cells i_1, i_2 at step n, the condition $x_1, x_2 > x_{i_1}^{(n-1)} - x_{i_2}^{(n-1)}$ (see Def.18) and single-view recognition with dual support, guarantees that the target map entropy at step n is smaller than it is at step $n-1$.

Any termination condition based on probability thresholding (*e.g.*, Def.7), requires a decreasing target map entropy. The above theorem quantifies a set of sufficient properties of the recognition algorithm, under which, multiple-view localization leads to a decreasing entropy and therefore, after a certain number of steps, a smaller target map entropy than that of a single-view. Theorem 3 lists all possible target map behaviours under the boundary constraints. If we also assume good single-view recognition and that no correspondence errors exist, Theorem 3 defines a set of sufficient properties of the single-view recognition algorithm so that multiple-view recognition leads to a decreasing target map entropy and a smaller bias and variance in the target's localization at each step. Without the boundary constraints, we have no guaranty of a decreasing entropy. Under good single-view recognition and a uniform target map prior, flipped cells are the result of correspondence errors, implying a possible increased target map entropy and bias in the target localization. Theorem 3 shows that without good single-view recognition, it is possible to have a decreasing target map entropy and an increasing bias in the estimated target position, exemplifying the difficulty of the problem.

Case (i) shows that if $x_{i_1}^{(n-1)}$ is the maximum probability amongst all cells at step $n-1$, the entropy decreases at the next step. It also shows that as the probabilities of the other cells relative to $x_{i_1}^{(n-1)}$ decrease, or as the relative weights $r_i^{(n)}$ for smaller probabilities increase (by decreasing their respective probabilities from step $n-1$ to step n, more than the other cells), it becomes more likely that the entropy will decrease in the next step. Case (ii) shows that a localization threshold of over $\frac{1}{2}$ easily leads to biased results under dead-reckoning or correspondence errors. Case (iv) is applicable when the correspondence errors increase and shows that more stringent requirements on the recognition algorithm can compensate for such errors and guarantee a decrease in the entropy (by requiring dual support and $x_1, x_2 > x_{i_1}^{(n-1)} - x_{i_2}^{(n-1)}$). No such requirement is needed in case (iii), which assumes that no flipped cells exist.

4.2. Proof of Theorem 2

Let $p(\mu_{v_n}|\neg c_j) \triangleq \frac{\sum_{i\neq j} p(\mu_{v_n}|c_i)p(c_i|\mu_{v_{n-1}},...,\mu_{v_1})}{\sum_{i\neq j} p(c_i|\mu_{v_{n-1}},...,\mu_{v_1})}$ (we assume a non-zero denominator). Since we only have dead-reckoning errors in (i), $\exists \hat{j}_t \in \hat{V}(v_n)$ such that $RT(i_t) = \hat{j}_t$. Thus $p(\mu_{v_n}|c_{\hat{j}_t}) > p(\mu_{v_n}|\neg c_{\hat{j}_t})$ regardless of $p(c_i|\mu_{v_{n-1}},...,\mu_{v_1})$ $\forall i \neq \hat{j}_t$, because if $p(\mu_{v_n}|c_{\hat{j}_t}) \geq p(\mu_{v_n}|\neg c_{\hat{j}_t})$ but not $p(\mu_{v_n}|c_{\hat{j}_t}) > p(\mu_{v_n}|\neg c_{\hat{j}_t})$ for all target map distributions at step $n-1$, there exists a cell $j \neq \hat{j}_t$ such that $p(\mu_{v_n}|c_j) = p(\mu_{v_n}|c_{\hat{j}_t})$ and thus $p(\mu_{v_n}|c_j) \geq p(\mu_{v_n}|\neg c_j)$ for all target maps, contradicting Def.16. Thus $\hat{j}_t = \arg\max_{j\in\hat{V}(v_n)} p(\mu_{v_n}|c_j) = \arg\max_{j\in\hat{V}(v_n)} E(Y_j^{(n)})$ (because of the uniform prior). Thus $\max_{j\in\hat{V}(v_n)} E(Y_j^{(n)}) = E(Y_{RT(i_t)}^{(n)})$ and since we have assumed to know the values of v_n, μ_{v_n}, any change in the dead-reckoning errors is equivalent to a change to the inertial coordinate frame and potentially to the label $RT(i_t)$ assigned to the structure represented by cell i_t, which does not affect $E(Y_{RT(i_t)}^{(n)})$, thus proving (i). Notice that $E(X_i^{(n)}) \leq E(X_i^{(n-1)}) \Leftrightarrow \frac{p(\mu_{v_n}|c_i)}{\sum_{a\in\{c_i,\neg c_i\}} p(a|\mu_{v_{n-1}},...,\mu_{v_1})p(\mu_{v_n}|a)} \leq 1$ which in conjunction with Lemma 2 below, proves (ii). The proof of case (iii) is similar to that of case (ii) and we leave it as an exercise. Case (iv) follows trivially from case (iii). Case (v) follows since $E(X_{\hat{i}_t}^{(n-1)})$ can be arbitrarily small and $E(X_{\hat{i}_t}^{(n)})$ is proportional to $E(X_{\hat{i}_t}^{(n-1)})$.

Lemma 2. *Let $g(x,\alpha,\beta) = \frac{\alpha}{\alpha x + \beta(1-x)}$ with $0 \leq \alpha, \beta, x \leq 1$ such that $\alpha x + \beta(1-x) \neq 0$. Then $g(x,\alpha,\beta) \leq 1$ iff $\beta > \alpha$ or $x = 1$ or $\alpha = \beta$.*

Proof. Notice that $g(x,\alpha,\beta) \leq 1 \Leftrightarrow \alpha - \beta \leq (\alpha-\beta)x$. If $\alpha < \beta$, then $g(x,\alpha,\beta)$ holds iff $x \leq 1$ which we know is always true. If $\alpha > \beta$, then $g(x,\alpha,\beta)$ holds iff $x = 1$. If $\alpha = \beta$, then $g(x,\alpha,\beta) = 1$ which proves the lemma. □

4.3. Proof of Theorem 3

To simplify certain arguments, we assume that no cell ever takes a value of zero. Let $X_i^{(n)}$ denote a Bernoulli random variable with probability of success $x_i^{(n)} \triangleq p(c_i|\mu_{v_n},...,\mu_{v_1})$. By Lemma 3, the boundary constraint assumption of Theorem 3 is equivalent to $Var(X_i^{(n)}) < Var(X_i^{(n-1)})$ $\forall i \in C$. Since the variance of a Bernoulli(p) random variable is equal to $p(1-p)$, it is maximized at $p = \frac{1}{2}$ and it is also symmetric around $p = \frac{1}{2}$, which implies that when the variance of $X_i^{(n)}$ has decreased, $|x_i^{(n)} - \frac{1}{2}| > |x_i^{(n-1)} - \frac{1}{2}|$. Since $Var(X_i^{(n)}) < Var(X_i^{(n-1)})$ for all cells i, there exists exactly one cell i_1 at step n with $x_{i_1}^{(n)} > \frac{1}{2}$, since otherwise, the variance of all cells could not have decreased and maintained a sum of one across all target map cells. This proves the first half of Theorem 3(i). One of the following conditions must hold at each step n:

(1): $\forall i$, $x_i^{(n-1)} \leq \frac{1}{2}$ and there exists exactly one cell i_1 that satisfies $x_{i_1}^{(n)} > \frac{1}{2}$.

(2): There exist two cells i_1, i_2 such that $x_{i_1}^{(n-1)} > \frac{1}{2}$, $x_{i_2}^{(n-1)} < \frac{1}{2}$, $x_{i_1}^{(n)} > \frac{1}{2}$, $x_{i_2}^{(n-1)} < \frac{1}{2}$.

(3): $x_{i_1}^{(n-1)} > \frac{1}{2}, x_{i_2}^{(n-1)} < \frac{1}{2}, x_{i_1}^{(n)} < \frac{1}{2}, x_{i_2}^{(n)} > \frac{1}{2}$.

Assume condition (1) applies. We now prove the second half of Theorem 3(i). For notational simplicity we index the $|C| - 1$ cells that are not equal to i_1 by the set $\{1, 2, ..., |C| - 1\}$. Let $g(p) \triangleq -p \lg(p)$. We want to show that $g(x_{i_1}^{(n-1)}) + \sum_{i=1}^{|C|-1} g(x_i^{(n-1)}) > g(x_{i_1}^{(n)}) + \sum_{i=1}^{|C|-1} g(x_i^{(n)})$ or equivalently $\sum_{i=1}^{|C|-1} \frac{g(x_i^{(n-1)}) - g(x_i^{(n-1)} - d_i^{(n)})}{x^{(n)}} > \frac{g(x_{i_1}^{(n-1)} + x^{(n)}) - g(x_{i_1}^{(n-1)})}{x^{(n)}}$ where $x^{(n)} \triangleq \sum_{i=1}^{|C|-1} d_i^{(n)}$. Notice that because of the boundary constraint and Lemma 3, $d_i^{(n)} > 0$ for $i \neq i_1$ and $x_i^{(n)} = x_{i_1}^{(n-1)} + x^{(n)}$ since the target map cells have to sum to one at step n. By the Mean Value Theorem, for each $i \in \{1, ..., |C| - 1\}$, $\exists z_i \in [x_i^{(n-1)} - d_i^{(n-1)}, x_i^{(n-1)}]$ such that $g(x_i^{(n-1)}) - g(x_i^{(n-1)} - d_i^{(n)}) = d_i^{(n)} g'(z_i)$ and $\exists z \in [x_{i_1}^{(n-1)}, x_{i_1}^{(n-1)} + x^{(n)}]$ such that $g(x_{i_1}^{(n-1)} + x^{(n)}) - g(x_{i_1}^{(n-1)}) = x^{(n)} g'(z)$. Notice that $\sum_{i=1}^{|C|-1} r_{i,i_1}^{(n)} = 1$ and $g'(p) = -\frac{\log(p)}{\log(2)} - \frac{1}{\log(2)}$. This in turn implies that $\sum_{i=1}^{|C|-1} r_{i,i_1}^{(n)} g'(z_i) > g'(z)$ and the entropy decreases if and only if $\prod_{i=1}^{|C|-1} z_i^{r_{i,i_1}^{(n)}} < z$. But since $\prod_{i=1}^{|C|-1} z_i^{r_{i,i_1}^{(n)}} \leq \prod_{i=1}^{|C|-1} (x_i^{(n-1)})^{r_{i,i_1}^{(n)}}$ and $x_{i_1}^{(n-1)} \leq z$, a sufficient condition for a decrease in the entropy is $x_{i_1}^{(n-1)} > \prod_{i=1}^{|C|-1} (x_{i,i_1}^{(n-1)})^{r_{i,i_1}^{(n)}}$. This proves (i).

The proof of part (ii) of the theorem follows, since if $x_i^{(n)} \leq \frac{1}{2}$ for all cells $i \in C$, then the probability of cell i_1 has decreased at step n ($x_{i_1}^{(n)} \leq \frac{1}{2} < x_{i_1}^{(n-1)}$) and for at least one cell i_2, $x_{i_2}^{(n-1)} < x_{i_2}^{(n)} \leq \frac{1}{2}$ so that all cell probabilities sum to one at step n. But this contradicts the monotonically decreasing variances implied by Lemma 3, proving (ii).

If condition (2) holds, by a recursive application of Lemma 4 (by setting $\gamma = \frac{1}{2}$), we see that $-\sum_{i \in C} x_i^{(n)} \lg(x_i^{(n)}) < -\sum_{i \in C} x_i^{(n-1)} \lg(x_i^{(n-1)})$ as desired. This proves part (iii) of the theorem.

For the proof of part (iv) of the theorem, condition (3) applies. Notice that $g(p)$ is monotonically increasing on $(0, \frac{1}{e}]$ and monotonically decreasing on $(\frac{1}{2}, 1]$. Since we have assumed $x_1, x_2 > x_{i_1}^{(n-1)} - x_{i_2}^{(n-1)}$, it suffices to show that $g(x_{i_1}^{(n-1)}) + g(x_{i_2}^{(n-1)}) > g(x_{i_1}^{(n)}) + g(x_{i_2}^{(n)})$ (since the probabilities of all cells $i \neq i_1, i_2$ have decreased and we assume dual support). Equivalently, we want to show that $g(x_{i_1}^{(n-1)}) + g(x_{i_2}^{(n-1)}) > g(x_{i_1}^{(n-1)} - x_1) + g(x_{i_2}^{(n-1)} + x_2)$. But since $x_{i_2}^{(n-1)} > x_{i_1}^{(n-1)} - x_1$, $x_{i_1}^{(n-1)} < x_{i_2}^{(n-1)} + x_2$ and we have assumed dual support ($x_{i_1}^{(n-1)} > \frac{1}{2}$, $x_{i_2}^{(n-1)} < \frac{1}{e}$), we have proven part (iv) of the theorem.

Lemma 3. *$Var(X_i^{(n)}) < Var(X_i^{(n-1)})$ if and only if $p(\mu_{v_n}|c_i) < p(\mu_{v_n}|\neg c_i)$ and*

$$p(c_i|\mu_{v_{n-1}}, ..., \mu_{v_1}) < \frac{p(\mu_{v_n}|\neg c_i) - \sqrt{p(\mu_{v_n}|c_i)p(\mu_{v_n}|\neg c_i)}}{p(\mu_{v_n}|\neg c_i) - p(\mu_{v_n}|c_i)}$$

or $p(\mu_{v_n}|c_i) > p(\mu_{v_n}|\neg c_i)$ and

$$p(c_i|\mu_{v_{n-1}}, ..., \mu_{v_1}) > \frac{p(\mu_{v_n}|\neg c_i) - \sqrt{p(\mu_{v_n}|c_i)p(\mu_{v_n}|\neg c_i)}}{p(\mu_{v_n}|\neg c_i) - p(\mu_{v_n}|c_i)}$$

Proof. For notational simplicity, let $\alpha = p(\mu_{v_n}|c_i)$, $\beta = p(\mu_{v_n}|\neg c_i)$ and $x = p(c_i|\mu_{v_{n-1}}, ..., \mu_{v_1})$. Notice that $Var(X_i^{(n)}) = E(X_i^{(n)})(1 - E(X_i^{(n)}))$, $E(X_i^{(n)}) = x\frac{\alpha}{\alpha x + \beta(1-x)}$ and $1 - E(X_i^{(n)}) = (1-x)\frac{\beta}{\alpha x + \beta(1-x)}$. Thus,

$$\frac{Var(X_i^{(n)})}{Var(X_i^{(n-1)})} < 1 \Leftrightarrow \frac{\alpha\beta}{(\alpha x + \beta(1-x))^2} < 1 \Leftrightarrow$$
$$0 < (\alpha - \beta)^2 x^2 + 2\beta(\alpha - \beta)x + \beta(\beta - \alpha) \triangleq g(x). \quad (3)$$

The zeros of $g(x)$ are $x = \frac{\beta \pm \sqrt{\alpha\beta}}{\beta - \alpha}$. Notice that the zeros of $g(x)$ are independent of changes in α, β that retain the ratio of α to β. Since $g(x)$ is a quadratic function of x, we can determine the range of values that satisfy (3): If $\alpha < \beta$, $g(x)$ is concave up, $\frac{\beta + \sqrt{\alpha\beta}}{\beta - \alpha} \geq 1$, and $\frac{\beta - \sqrt{\alpha\beta}}{\beta - \alpha} \leq \frac{\beta + \sqrt{\alpha\beta}}{\beta - \alpha}$ which implies that (3) is satisfied iff $x < \frac{\beta - \sqrt{\alpha\beta}}{\beta - \alpha}$ (see the graph of $f_2(\alpha)$ in Fig.(1)). If $\alpha > \beta$, $g(x)$ is again concave up, $\frac{\beta + \sqrt{\alpha\beta}}{\beta - \alpha} \leq 0$, and $\frac{\beta + \sqrt{\alpha\beta}}{\beta - \alpha} \leq \frac{\beta - \sqrt{\alpha\beta}}{\beta - \alpha}$ which implies that (3) is satisfied iff $x > \frac{\beta - \sqrt{\alpha\beta}}{\beta - \alpha}$ (see the graph of $f_1(\beta)$ in Fig.(1)). The lemma shows that there exist a set of achievable probability values based on a relationship between the quality of the discriminative and single-view generative probabilities that is sufficient to decrease the variance of each cell's Bernoulli distribution. In conjunction with Theorem 3, it demonstrates the strong relationship between each cell's variance and the target map entropy. □

Lemma 4. *Let $g(p) \triangleq -p \lg(p)$, where $0 < p \leq 1$. Let $\gamma \in (0, 1)$, $0 < \alpha < \gamma < \beta \leq 1$, $0 < x < \alpha$ and $0 < x \leq 1 - \beta$. Then $g(\alpha) + g(\beta) > g(\alpha - x) + g(\beta + x)$.*

Proof. Notice that $g(\alpha) + g(\beta) > g(\alpha - x) + g(\beta + x) \Leftrightarrow g(\alpha) - g(\alpha - x) > g(\beta + x) - g(\beta) \Leftrightarrow \frac{g(\alpha) - g(\alpha - x)}{x} > \frac{g(\beta + x) - g(\beta)}{x}$. By the Mean Value Theorem, there exist $\alpha' \in [\alpha - x, \alpha]$ and $\beta' \in [\beta, \beta + x]$ such that $g'(\alpha') = \frac{g(\alpha) - g(\alpha - x)}{x}$ and $g'(\beta') = \frac{g(\beta + x) - g(\beta)}{x}$. But since $g'(p) = -\frac{\log(p)}{\log(2)} - \frac{1}{\log(2)}$ is a decreasing function and $\forall \alpha'' \in [\alpha - x, \alpha]\ \forall \beta'' \in [\beta, \beta + x]$, $g'(\alpha'') > g'(\beta'')$, we have proven the lemma. □

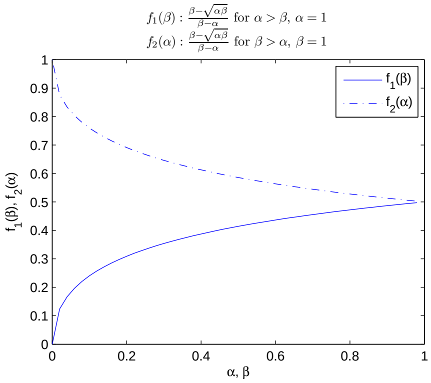

Figure 1. Graphs showing on the y-axis the zeros of $g(x)$ for various ratios of α, β (see Ineq.(3) in Lemma 3).

5. Discussion

A number of optimization algorithms for navigation, mapping and next-view-planning have been suggested, based on POMDPs [11], Bayesian methods [12, 13], heuristics [5, 8, 20] and greedy algorithms [15, 21] amongst others. The arguments in Sec. 3 suggest what kind of policies would lead to efficient and reliable solutions for the components of active object detection and localization systems that deal with sensor control for recognition. These include next-view-planners that use efficient approximation algorithms, or algorithms based on greedy and dynamic programming solutions to the Knapsack problem [7, 21], suggesting that a mixture of specialized optimizers, rather than a single kind of optimization, could lead to more efficient solutions, without a significant decrease in reliabilty.

Theorems 2,3 suggest that single-view localization, where the updated target map is only used to guide the where-to-look-next policy, can lead to fewer false positive/negative detections, at the expense of greater localization bias when dead-reckoning errors occur. Alternatively, if we have some prior knowledge about the expected maximum dead reckoning error—typically the main source of correspondence errors—, we can define appropriate dimensions for cell i_t, such that the target's centroid always falls inside cell i_t. For example, an adaptive multiscale target map approach could be used. At each step, we could adjust the scale of the cells close to the expected target position, based on the expected dead-reckoning errors, in order to guarantee a monotonically decreasing target map entropy. This would make a termination condition based on probability thresholding (*e.g.*, Def.7) more reliable under modest dead-reckoning errors, despite potentially increased target localization bias. At that point, target re-localization could take place within this region, to refine the target position.

6. Conclusions

We have proven that the active object localization problem and a number of its variants, are NP-Hard or NP-Complete. We have studied the tradeoffs of localizing vs. detecting a target object under single-view and multiple-view recognition schemes. We have shown that a number of bias/variance/entropy relationships and tradeoffs emerge under single-view and multi-view localization and detection schemes, that depend on the quality of the recognition algorithm and the magnitudes of the correspondence or dead-reckoning errors. The results motivated a set of properties for active detection and localization algorithms.

References

[1] Aristotle. *Περί Ψυχής (On the Soul)*. 350 B.C. 1

[2] R. Bajcsy. Active perception vs. passive perception. In *IEEE Workshop on Computer Vision Representation and Control*, Bellaire, Michigan, 1985. 1

[3] F. Callari and F. Ferrie. Active recognition: Looking for differences. *Int. J. Comput. Vision*, 43(3):189–204, 2001. 1

[4] S. Dickinson, H. Christensen, J. Tsotsos, and G. Olofsson. Active object recognition integrating attention and viewpoint control. *Comput. Vis. Image Und.*, 67(3):239–260, 1997. 1

[5] S. Dickinson, D. Wilkes, and J. Tsotsos. A computational model of view degeneracy. *IEEE Trans. Patt. Anal. Mach. Intell.*, 21(8):673–689, August 1999. 1, 8

[6] S. Ekvall, P. Jensfelt, and D. Kragic. Integrating active mobile robot object recognition and SLAM in natural environments. In *Proc. Intelligent Robots and Systems*, 2006. 1

[7] M. R. Garey and D. S. Johnson. *Computers and Intractability: A guide to the Theory of NP-Completeness*. W.H. Freeman and Company, 1979. 2, 3, 8

[8] T. Garvey. Perceptual strategies for purposive vision. Technical report, Nr. 117, SRI Int'l., 1976. 1, 8

[9] W. E. L. Grimson. The combinatorics of heuristic search termination for object recognition in cluttered environments. *IEEE Trans. Patt. Anal. Mach. Intell.*, 13:920–935, 1991. 1

[10] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman and Company, 1982. 1

[11] D. Meger, P. Forssen, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. Little, and D. Lowe. Curious George: An attentive semantic robot. In *Proc. Robot. Auton. Syst.*, 2008. 1, 8

[12] R. D. Rimey and C. M. Brown. Control of selective perception using bayes nets and decision theory. *Int. J. Comput. Vision*, 12(2/3):173–207, 1994. 1, 8

[13] S. D. Roy, S. Chaudhury, and S. Banerjee. Isolated 3D object recognition through next view planning. *IEEE Trans. Syst. Man Cybern. Part A Syst. Humans*, 30(1):67–76, 2000. 1, 8

[14] W. Rudin. *Principles of Mathematical Analysis*. McGraw Hill, 1976. 2

[15] F. Saidi, O. Stasse, K. Yokoi, and F. Kanehiro. Online object search with a humanoid robot. In *Proc. Intelligent Robots and Systems*, 2007. 1, 8

[16] B. Schiele and J. Crowley. Transinformation for active object recognition. In *Proc. Int. Conf. on Computer Vision*, 1998. 1

[17] J. Tsotsos, Y. Liu, J. Martinez-Trujillo, M. Pomplun, E. Simine, and K. Zhou. Attending to visual motion. *Comput. Vis. Image Und.*, 100(1-2):3–40, 2005. 1

[18] J. K. Tsotsos. Analyzing vision at the complexity level. *Behav. Brain Sci.*, 13-3:423–445, 1990. 1

[19] J. K. Tsotsos. On the relative complexity of active vs. passive visual search. *Int. J. Comput. Vision*, 7(2):127–141, 1992. 1

[20] L. E. Wixson and D. H. Ballard. Using intermediate objects to improve the efficiency of visual search. *Int. J. Comput. Vision*, 12(2/3):209–230, 1994. 1, 8

[21] Y. Ye. *Sensor Planning in 3D Object Search*. PhD thesis, University of Toronto, 1997. 1, 3, 8