Sensor Planning in 3D Object Search: its Formulation and Complexity

Yiming Ye and **John K. Tsotsos**

Department of Computer Science

University of Toronto

Toronto, Ontario, Canada M5S 1A4

yiming@vis.toronto.edu tsotsos@vis.toronto.edu

1 Introduction

The research described in this paper conforms to the complexity level analysis of the sensor planning task for object search. Complexity considerations are commonplace in the biological and computational vision literature. For example, Tsotsos [8] shows that the general problem of visual search (search for a target within an image) is computationally intractable in a formal, complexity-theoretic sense. He [9] also ties the concept of active perception to attentive processing in general and to his complexity level analysis of visual search and proves that active unbounded visual search is NP-Complete. Kirousis and Papadimitriou [4] show that the problem of polyhedral scene labeling is inherently NP-Complete. Many other vision researchers ([3],[6], etc.) routinely provide an analysis of the complexity of their proposed algorithms. Complexity level analysis of robotics and vision problems is important because it can reveal basic insights into the structure of the problem and delimit the space of permissible solutions in a formal and theoretical fashion.

Object search is the task of searching for a given $3D$ object in a given $3D$ environment. Sensor planning for object search refers to the task of how to select the sensing parameters so as to bring the target into the field of view of the sensor. This task is very important if a robot wants to interact intelligently and effectively with its environment. Connell [1] constructs a robot that roams an area searching for and collecting soda cans. The planning is very simple since the robot just follows the walls of the room and the sensor only searches the area immediately in front of the robot. Rimey and Brown [7] use composite Bayes net and utility decision rule to plan the sensor in their task-oriented system TEA. The indirect search mechanism proposed by Garvey [2] is to first direct the sensor to search for an "intermediate" object that commonly participates in a spatial relationship with the target and then direct the sensor to examine the restricted region specified by this relationship. Wixson and Ballard [10] present a mathematical model of search efficiency and predict that indirect search can improve efficiency in many situations. It is interesting to note that the operational research community has done a lot of research on optimal search [5]. Their purpose is to determine how to allocate effort to search for a target, such as a lost submarine in the ocean or an oil field within a certain region. Although the results are elegant and beautiful in a mathematical sense, they can not be directly applied here because the searcher model is too abstract and general and there is no sensor planning involved in their approach.

There is no previous research within the computer vision community that attempts to formalize the sensor planning task for object search in general and to analyze this problem at the complexity level. This paper is an attempt along this direction. By combining the **probability distribution** of the target and the **detecting ability** of the recognition algorithms, we formulate the sensor planning problem for object search, discuss several properties of this task, and prove that this task is NP-Complete. The theoretical result provided in this paper has been used as a guideline in designing the practical sensing strategies (see [11] for detail).

2 Problem Formulation

Although it is important to examine different aspects of object search individually and in some degree of isolation, it is even more important to study their relationship and how to integrate them into a whole search system.

The search region Ω can be in any form and it is assumed that we know the boundary of Ω exactly but we do not know its internal configuration. In practice, we will tessellate the region Ω into a series of little elements c_i (here, we assume that those elements are in the form of little cubes), $\Omega = \bigcup_{i=1}^{n} c_i$ and $c_i \bigcap c_j = 0$.

The searcher is a mobile platform equipped with a camera that can pan, tilt and zoom. The state of the searcher is uniquely determined by 7 parameters $(x_c, y_c, z_c, p, t, w, h)$. (x_c, y_c, z_c) is the position of the camera center (the starting point of the camera viewing axis), (p, t) is the direction of the camera viewing axis (p is the amount of pan, t is the amount of tilt). w, h are the width and height of the solid viewing angle of the camera. (x_c, y_c, z_c) can be adjusted by moving the mobile platform. (p, t) can be adjusted by the motors on the robotics head. w, h can be adjusted by the zoom lens of the camera. Only finite number of platform positions is allowed.

An operation $\mathbf{f} = \mathbf{f}(x_c, y_c, z_c, p, t, w, h, a)$ is an action of the searcher within the region Ω, where a is a recognition algorithm. An operation $\mathbf{f}$ entails: take a **perspective** projection image according to the current camera configuration and then search the image using the given recognition algorithm. The number of total different operations is big, but is not infinite. This number is determined by hardware properties of the mobile platform, the robotics head, the zoom camera, and the available recognition algorithms [11].

The target distribution can be specified by a probability distribution function $\mathbf{p}$. $\mathbf{p}(c_i, t)$ gives the probability that the center of the target is within cube c_i at time t.

The detection function on Ω is a function $\mathbf{b}$, such that $\mathbf{b}(c_i, \mathbf{f})$ gives the conditional probability of detecting the target given that the center of the target is located within c_i and the operation is $\mathbf{f}$. For any operation, if the projection of the center of the cube c_i is outside the image, we assume $\mathbf{b}(c_i, \mathbf{f}) = 0$; if it is too far from the camera or too near to the camera, we also have $\mathbf{b}(c_i, \mathbf{f}) = 0$. In general [11], $\mathbf{b}(c_i, \mathbf{f})$ is determined by various factors, such as intensity, occlusion, and orientation etc.. It is obvious that the probability of detecting the target by applying action $\mathbf{f}$ is given by $P(\mathbf{f}) = \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}})\mathbf{b}(c_i, \mathbf{f})$. Where $t_{\mathbf{f}}$ is the time just before $\mathbf{f}$ is applied.

The reason that the term $t_{\mathbf{f}}$ is introduced in the calculation of $P(\mathbf{f})$ is because that the probability distribution need to be updated whenever an action is failed. Here Bayes' formula is used to incorporate the new recognition results into the old distribution. Let α_i be the event that the center of the target is in cube c_i, α_o be the event that the center of the target is outside the region, let β be the event that after applying a recognition action, the recognizer successfully detects the target. Since the above events $\alpha_1, \ldots, \alpha_n, \alpha_o$ are mutually complementary and exclusive, we can get the following formula

$$P(\alpha_i \mid \neg\beta) = \frac{P(\alpha_i)P(\neg\beta \mid \alpha_i)}{P(\alpha_o)P(\neg\beta \mid \alpha_o) + \sum_{j=1}^{n} P(\alpha_j)P(\neg\beta \mid \alpha_j)}, \quad i = 1, \ldots, n, o \tag{1}$$

Thus we have the following updating rule (Note: $t_{\mathbf{f}+}$ refers to the time just after the action $\mathbf{f}$ is applied):

$$\mathbf{p}(c_i, t_{\mathbf{f}+}) \leftarrow \frac{\mathbf{p}(c_i, t_{\mathbf{f}})(1-\mathbf{b}(c_i, \mathbf{f}))}{\mathbf{p}(c_o, t_{\mathbf{f}}) + \sum_{j=1}^{n} \mathbf{p}(c_j, t_{\mathbf{f}})(1-\mathbf{b}(c_j, \mathbf{f}))}, \quad i = 1, \ldots, n, o \tag{2}$$

Since $\mathbf{p}(c_o, t_{\mathbf{f}}) + \sum_{j=1}^{n} \mathbf{p}(c_j, t_{\mathbf{f}}) = 1$ and $P(\mathbf{f}) = \sum_{j=1}^{n} \mathbf{p}(c_j, t_{\mathbf{f}})\mathbf{b}(c_j, \mathbf{f})$, the updating rule becomes

$$\mathbf{p}(c_i, t_{\mathbf{f}+}) \leftarrow \frac{\mathbf{p}(c_i, t_{\mathbf{f}})(1-\mathbf{b}(c_i, \mathbf{f}))}{1 - P(\mathbf{f})}, \quad i = 1, \ldots, n, o \tag{3}$$

The cost function $\mathbf{t}(\mathbf{f})$ gives the time required to execute action $\mathbf{f}$. This include the time needed to adjust the camera configuration to the status specified by $\mathbf{f}$ and the time needed to take a picture under current configuration and run the recognition algorithm. We assume that the cost for a given operation is fixed and it is not influenced by the previous actions.

Let $\mathbf{O}_{\mathbf{\Omega}}$ be the set of all possible operations that can be applied. The effort allocation $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_q\}$ gives the ordered set of operations applied in the search, where $\mathbf{f}_i \in \mathbf{O}_{\mathbf{\Omega}}$. It is clear that the probability of detecting the target by this allocation is:

$$\begin{aligned} P[\mathbf{F}] \;=\; & \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}_1})\mathbf{b}(c_i, \mathbf{f}_1) + [1 - \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}_1})\mathbf{b}(c_i, \mathbf{f}_1)] \times [\sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}_2})\mathbf{b}(c_i, \mathbf{f}_2)] \\ & + \ldots + \{ \prod_{j=1}^{q-1} [1 - \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}_j})\mathbf{b}(c_i, \mathbf{f}_j)] \} \times [\sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}_q})\mathbf{b}(c_i, \mathbf{f}_q)] \end{aligned} \tag{4}$$

and the total time for applying this allocation is (following [9]):

$$T[\mathbf{F}] = \sum_{1 \le i \le q} \mathbf{t}(\mathbf{f}_i) \tag{5}$$

Suppose K is the total time that can be allowed in the search, then the task of sensor planning for object search can be defined as finding an allocation $\mathbf{F} \subset \mathbf{O}_{\mathbf{\Omega}}$, which satisfies $T(\mathbf{F}) \le K$ and maximizes $P[\mathbf{F}]$.

3 Some properties of the object search process

We list some properties of the sensor planning task in this section, proofs are omitted.

Suppose $\Omega = \bigcup_{i=1}^{n} c_i$, $\mathbf{O}_{\mathbf{\Omega}} = \{\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_m^*\}$. For any operation $\mathbf{f} \in \mathbf{O}_{\mathbf{\Omega}}$, we define its influence range as $\Omega(\mathbf{f}) = \{c \mid \mathbf{b}(c, \mathbf{f}) \neq 0\}$. Its complement form is: $\overline{\Omega(\mathbf{f})} = \Omega - \{c \mid \mathbf{b}(c, \mathbf{f}) \neq 0\}$.

For any effort allocation $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_q \mid \mathbf{f}_i \in \mathbf{O}_{\mathbf{\Omega}}\}$. The initial probability distribution is denoted as $\mathbf{p}^{[0]}(c_1), \mathbf{p}^{[0]}(c_2), \ldots, \mathbf{p}^{[0]}(c_n), \mathbf{p}^{[0]}(c_o)$. After the application of the operation $\mathbf{f}_1$, the distribution is denoted by $\mathbf{p}^{[1]}(c_1), \mathbf{p}^{[1]}(c_2), \ldots, \mathbf{p}^{[1]}(c_n), \mathbf{p}^{[1]}(c_o)$. Generally, after the application of the operation $\mathbf{f}_i$, the distribution is denoted by $\mathbf{p}^{[i]}(c_1), \mathbf{p}^{[i]}(c_2), \ldots, \mathbf{p}^{[i]}(c_n), \mathbf{p}^{[i]}(c_o)$, where $1 \leq i \leq q$. Let $P(\mathbf{f}_i)$ represent the probability of detecting the target by applying the action $\mathbf{f}_i$ with respect to the allocation $\mathbf{F}$. Then of course we have $P(\mathbf{f}_i) = \sum_{j=1}^{n} \mathbf{p}^{[i-1]}(c_j)\mathbf{b}(c_j, \mathbf{f_i})$. Let $P^{[0]}(\mathbf{f}_i)$ represents the probability of detecting the target by applying the action $\mathbf{f}_i$ when there is no action been applied before, then we have $P^{[0]}(\mathbf{f}_i) = \sum_{j=1}^{n} \mathbf{p}^{[0]}(c_j)\mathbf{b}(c_j, \mathbf{f_i})$.

Lemma 1. For allocation $\mathbf{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_q \mid \mathbf{f}_i \in \mathbf{O_\Omega}\}$, we have

$$\mathbf{p}^{[k]}(c) = \frac{\mathbf{p}^{[0]}(c)t_1(c)t_2(c)\ldots t_k(c)}{(1-P(\mathbf{f}_1))(1-P(\mathbf{f}_2))\ldots,(1-P(\mathbf{f}_k))} \tag{6}$$

where

$$t_i(c) = \begin{cases} 1 - \mathbf{b}(c, \mathbf{f}_i) & \text{if } c \in \Omega(\mathbf{f}_i) \\ 1 & \text{otherwise} \end{cases}, i = 1, \ldots, k$$

Definition 1 For any group of actions $\{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$, we define:

$$\begin{aligned} Q_k^{[0]}(\mathbf{f}_1 \ldots \mathbf{f}_k) &= \Omega(\mathbf{f}_1)\bigcap\ldots\bigcap\Omega(\mathbf{f}_k) \\ &\ldots \\ Q_k^{[r]}(\mathbf{f}_1 \ldots \overline{\mathbf{f}_{i_1}} \ldots \overline{\mathbf{f}_{i_2}} \ldots \overline{\mathbf{f}_{i_r}} \ldots \mathbf{f}_k) &= \Omega(\mathbf{f}_1)\bigcap\ldots\bigcap\overline{\Omega(\mathbf{f}_{i_1})}\bigcap\ldots\bigcap\overline{\Omega(\mathbf{f}_{i_2})}\bigcap\ldots\bigcap\overline{\Omega(\mathbf{f}_{i_r})}\bigcap\ldots\bigcap\Omega(\mathbf{f}_k); \\ &\ldots \\ Q_k^{[k]}(\overline{\mathbf{f}_1\mathbf{f}_2}\ldots\overline{\mathbf{f}_k}) &= \overline{\Omega(\mathbf{f}_1)}\bigcap\ldots\bigcap\overline{\Omega(\mathbf{f}_k)}; \end{aligned}$$

Where $Q_k^{[r]}$ $(r \leq k)$ means that there are k sets $\Omega(\mathbf{f}_1)$, $\Omega(\mathbf{f}_2)$, ..., $\Omega(\mathbf{f}_k)$ that are taken into the consideration, r of them are in their complement form.

For a given k, it is easy to see that the intersection of any two different $Q_k^{[r]}$ is ϕ. The unification of all the possible $Q_k^{[r]}$ is Ω, as stated in **Lemma 2.**

Lemma 2. For the above defined $Q_k^{[r]}$, we have:

$$\bigcup_{r=0}^{k}\Big\{\bigcup_{1\leq i_1<i_2<\ldots<i_r\leq k} Q_k^{[r]}(\mathbf{f}_1 \ldots \overline{\mathbf{f}_{i_1}} \ldots \overline{\mathbf{f}_{i_2}} \ldots \overline{\mathbf{f}_{i_r}} \ldots \mathbf{f}_k)\Big\} = \Omega$$

Lemma 3 For allocation $\mathbf{F} = \{\mathbf{f}_1 \dots \mathbf{f}_q\}$, we have

$$
\begin{aligned}
P(\mathbf{f}_k) \;=\; & \frac{1}{(1-P(\mathbf{f}_1))\dots(1-P(\mathbf{f}_{k-1}))}\Big\{P^{[0]}(\mathbf{f}_k) \\
& +(-1)^1\sum_{i_1=1}^{k-1}\Big(\sum_{c\in\Omega(\mathbf{f}_{i_1})\bigcap\Omega(\mathbf{f}_k)}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_{i_1})\mathbf{b}(c,\mathbf{f}_k)\Big) \\
& +(-1)^2\sum_{1\le i_1<i_2\le k-1}\Big(\sum_{c\in\Omega(\mathbf{f}_{i_1})\bigcap\Omega(\mathbf{f}_{i_2})\bigcap\Omega(\mathbf{f}_k)}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_{i_1})\mathbf{b}(c,\mathbf{f}_{i_2})\mathbf{b}(c,\mathbf{f}_k)\Big) \\
& +(-1)^3\sum_{1\le i_1<i_2<i_3\le k-1}\Big(\sum_{c\in\Omega(\mathbf{f}_{i_1})\bigcap\Omega(\mathbf{f}_{i_2})\bigcap\Omega(\mathbf{f}_{i_3})\bigcap\Omega(\mathbf{f}_k)}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_{i_1})\mathbf{b}(c,\mathbf{f}_{i_2})\mathbf{b}(c,\mathbf{f}_{i_3})\mathbf{b}(c,\mathbf{f}_k)\Big) \\
& +\dots \\
& +(-1)^{k-1}\sum_{c\in\Omega(\mathbf{f}_1)\bigcap\dots\bigcap\Omega(\mathbf{f}_{k-1})\bigcap\Omega(\mathbf{f}_k)}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_1)\mathbf{b}(c,\mathbf{f}_2)\dots\mathbf{b}(c,\mathbf{f}_{k-1})\mathbf{b}(c,\mathbf{f}_k)\Big\}
\end{aligned}
$$

for $k = 2, \dots, q$.

Lemma 4 For the given allocation $\mathbf{F} = \{\mathbf{f_1}, \dots, \mathbf{f_q}\}$, we have

$$
\begin{aligned}
P(\mathbf{F}) \;=\; & \sum_{i_1=1}^{q}P^{[0]}(\mathbf{f}_i) \\
& +(-1)^{2+1}\sum_{1\le i_1<i_2\le q}\Big(\sum_{c\in\Omega(\mathbf{f}_{i_1})\bigcap\Omega(\mathbf{f}_{i_2})}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_{i_1})\mathbf{b}(c,\mathbf{f}_{i_2})\Big) \\
& \dots \\
& +(-1)^{r+1}\sum_{1\le i_1<i_2<\dots<i_r\le q}\Big(\sum_{c\in\Omega(\mathbf{f}_{i_1})\bigcap\Omega(\mathbf{f}_{i_2})\bigcap\dots\bigcap\Omega(\mathbf{f}_{i_r})}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_{i_1})\mathbf{b}(c,\mathbf{f}_{i_2})\dots\mathbf{b}(c,\mathbf{f}_{i_r})\Big) \\
& \dots \\
& +(-1)^{q+1}\Big(\sum_{c\in\Omega(\mathbf{f}_1)\bigcap\Omega(\mathbf{f}_2)\bigcap\dots\bigcap\Omega(\mathbf{f}_q)}\mathbf{p}^{[0]}(c)\mathbf{b}(c,\mathbf{f}_1)\dots\mathbf{b}(c,\mathbf{f}_q)\Big) \qquad (7)
\end{aligned}
$$

Lemma 5 For a given allocation $\mathbf{F} = \{\mathbf{f_1}, \dots, \mathbf{f_q}\}$, $P[\mathbf{F}]$ is not influenced by the order of the applied actions. That is to say: $P[\mathbf{F}]$ is symmetric about actions $\mathbf{f}_1, \dots, \mathbf{f}_q$.

Lemma 6 For an allocation $\mathbf{F} = \{\mathbf{f_1}, \ldots, \mathbf{f_q}\}$, if we restrict the available actions such that there is no cube belongs to any two action ranges: $\Omega(\mathbf{f}_j) \bigcap \Omega(\mathbf{f}_i) = \phi$, if $i \neq j$, then $P[\mathbf{F}]$ can be calculated by the following:

$$P(\mathbf{F}) = \sum_{i=1}^{q} P^{[0]}(\mathbf{f}_i) \tag{8}$$

4 The complexity of the object search task

In this section, we prove that the sensor planning task for object search is *NP-complete*. First, the maximization problem is changed into the equivalent decision problem.

INSTANCE: A search region $\Omega = \bigcup_{i=1}^{n} c_i$, a finite set $\mathbf{O_\Omega} = \{\mathbf{f_1}^*, \mathbf{f_2}^*, \ldots, \mathbf{f_m}^*\}$ (the set of all possible operations that can be applied), an expected probability function $P[\mathbf{F}']$ as described in (4) and a cost function $T[\mathbf{F}']$ as described in (5) for an effort allocation $\mathbf{F}'$. Where $\mathbf{F}' = \{\mathbf{f}_1, \ldots, \mathbf{f}_k \mid \mathbf{f}_i \in \mathbf{O_\Omega}\}$ is the ordered set of operations applied in the search. An environmental updating rule as described in (1), a cost constraint K and an expected probability goal M.

QUESTION: Is there an effort allocation $\mathbf{F}$ such that $T[\mathbf{F}] \leq K$ and $P[\mathbf{F}] \geq M$

It is obvious that the above problem belongs to NP. In the following, we use proof by restriction to prove that the object search task belongs to the NP-complete class.

Theorem 1 The sensor planning for object search task described in section 4.1 is NP-Complete.

Proof:

We prove the theorem by restriction.

Actually we already known that $KNAPSACK$ problem is NP-Complete. We will restrict the object search task into $KNAPSACK$ by allowing instances in which any two operations in $\mathbf{O_\Omega} = \{\mathbf{f_1}^*, \mathbf{f_2}^*, \ldots, \mathbf{f_m}^*\}$ do not have common action region. This is to say that $\Omega(\mathbf{f_i}) \bigcap \Omega(\mathbf{f_j}) = \phi$ for any $i \neq j$.

After this restriction, by combining **Lemma 5** and **Lemma 6**, the task becomes

INSTANCE: A search region $\Omega = \bigcup_{i=1}^{n} c_i$, a finite set $\mathbf{O}_{\Omega} = \{\mathbf{f_1}^*, \mathbf{f_2}^*, \ldots, \mathbf{f_m}^*\}$ (the set of all possible operations that can be applied), an expected probability function $P[\mathbf{F}'] = \sum_{i=1}^{k} P^{[0]}(\mathbf{f_i})$ (from Lemma 6) and a cost function $T[\mathbf{F}'] = \sum_{i=1}^{k} \mathbf{t}(\mathbf{f_i})$ for an effort allocation $\mathbf{F}'$. Where $\mathbf{F}' = \{\mathbf{f}_1, \ldots, \mathbf{f}_k \mid \mathbf{f}_i \in \mathbf{O}_{\Omega}\}$ is the set of operations applied in the search. (Note, here we do not have updating rule and the order of actions does not matter from **Lemma 5**). A cost constraint K and an expected probability goal M.

QUESTION: Is there an effort allocation $\mathbf{F}$ such that $T[\mathbf{F}] \leq K$ and $P[\mathbf{F}] \geq M$

Obviously the above problem is equivalent to the *KNAPSACK* problem. Thus, the task of sensor planning for object search is NP-Complete. □.

5 The simplified version: a one step look ahead problem

Since the sensor planning problem is NP-complete and usually the number of the available candidate actions are big, it is necessary to relax the original problem. Instead of looking for an algorithm that always generate an optimal solution, we will simply use heuristics that will generate a feasible solution for the original problem. Since the calculation of the detection probability for a given operation is time consuming, it is not practical to use complex strategies. Here we use greedy strategy. The greedy method suggests that one can device an algorithm which works in stages, considering one input at a time. At each stage, based on some optimization measure, the next candidate is selected and is included into the partial solution so far. Suppose we have already executed k ($k \geq 1$) actions $\mathbf{F}_k = \{\mathbf{f}_1, \ldots, \mathbf{f}_k\}$. We now want to find the next action $\mathbf{f}_{k+1}$ to execute, with the hope that our strategy of finding the next action may lead to an **approximate** solution for the object search task. Suppose $P[\mathbf{F}_q]$ is the expected probability of detecting the target by allocation $\mathbf{F}_q$. $T[\mathbf{F}_q]$ is the cost of allocation $\mathbf{F}_q$.

When we execute a next action $\mathbf{f}$, the effort allocation becomes $\mathbf{F}_{q+1} = \{\mathbf{f}_1, \ldots, \mathbf{f}_q, \mathbf{f}\}$. The expected probability of detecting the target is $P[\mathbf{F}_{q+1}] = P[\mathbf{F}_q] + \Delta_P(\mathbf{f})$, where $\Delta_P(\mathbf{f}) = \{\prod_{j=1}^{q}[1 - \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f_j}})\mathbf{b}(c_i, \mathbf{f_j})]\} \times [\sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f}})\mathbf{b}(c_i, \mathbf{f})]$. The total cost becomes $T[\mathbf{F}_{q+1}] = T[\mathbf{F}_q] + \Delta_T(\mathbf{f})$, where $\Delta_T(\mathbf{f}) = \mathbf{t}(\mathbf{f})$. Our strategy for selecting the next action is: the next action $\mathbf{f_{q+1}}$ should be selected that maximizes the term $\frac{\Delta_P(\mathbf{f})}{\Delta_T(\mathbf{f})}$ Since the term $\{\prod_{j=1}^{q}[1 - \sum_{i=1}^{n} \mathbf{p}(c_i, t_{\mathbf{f_j}})\mathbf{b}(c_i, \mathbf{f_j})]\}$ is fixed no matter what the next action $\mathbf{f}$ is selected, our strategy becomes: the next action $\mathbf{f_{q+1}}$ should be selected that maximizes the term

$$\mathbf{E}(\mathbf{f}) = \frac{\sum_{i=1}^{n} \mathbf{p}(c_i, t_\mathbf{f})\mathbf{b}(c_i, \mathbf{f})}{\mathbf{t}(\mathbf{f})}$$

In some situations, the simple greedy strategy may even generate the optimal answers as stated in Lemma 7.

Lemma 7: Suppose the available operations $\mathbf{O_\Omega} = \{\mathbf{f_1}, \mathbf{f_2}, \ldots, \mathbf{f_m}\}$ satisfy:

(1) $\mathbf{t}(\mathbf{f_i}) = \mathbf{t}(\mathbf{f_j})$, $1 \leq i, j \leq m$.

(2) $\Omega(\mathbf{f}_j) \bigcap \Omega(\mathbf{f}_i) = \phi$, $1 \leq i, j \leq m, i \neq j$.

Then, the above greedy strategy generates the optimal answer.

6 Discussion

We believe that it is very important to study the sensor planning problem of object search from a theoretical point of view. Only by this way can we get a better understanding of the problem and design an efficient and powerful algorithm in practice. This paper formalized the sensor planning problem, pointed out several properties of this task and analyzed the complexity of this task.

The result presented in this paper has been used as a guideline in designing the practical sensor planning system (see [11] for detail). According to the properties of the image formation process, we decompose the huge space of all possible sensing actions into a small number of actions that must be tried. These small number of actions have the property that, in most situations, the intersection of the influence ranges of any two of them is an empty set. The greedy strategy described in section 5 is used in the action selection process. Many of the properties discussed in this paper are applied in the implementation. Our sensor planning system has been extensively tested by simulation experiments and real experiments. The experimental results are satisfactory.

Acknowledgment: The second author is the CP-Unitel Fellow of the Canadian Institute for Advanced Research. The first author is grateful to Dr. D. Wilkes, Dr. P Jasiobedzki, and Mr. V. Lee for their comments.

References

[1] Connel. *An Artificial Creature*. PhD thesis, AI Lab, MIT, 1989.

[2] T. D. Garvey. Perceptual strategies for purposive vision. Technical Report Technical Note 117, SRI International, 1976.

[3] W. Grimson. The combinatorics of local constraints in model-based recognition and localization from sparse data. *Journal of the Association for Computing Machinery*, 33(4):658–686, 1986.

[4] L. Kirousis and C. Papadimitriou. The complexity of recognizing polyhedral scenes. In *26 Annual Symposium on Foundations of Computer Science*, Portland, Ore., 1985.

[5] B. O. Koopman. *Search and Screen: general principles with historical applications*. Pergaman Press, Elmsford, N.Y, 1980.

[6] W. H. Plantinga and C. R. Dyer. An algorithm for constructing the aspect graph. In *27 Annual Symposium on Foundations of Computer Science*, pages 123–131, Toronto, Ontario, 1986.

[7] R. Rimey and C. Brown. Where to look next using a bayes net: Incorporating geometric relations. In *Second European Conference on Computer Vision*, pages 542–550, Santa Margherita Ligure, Italy, May 1992.

[8] J. Tsotsos. Analyzing vision at the complexity level. *The behavioral and brain science*, 13:423–469, 1990.

[9] J. Tsotsos. Active verses passive perception, which is more efficient? *International Journal of Computer Vision*, 7(2), 1992.

[10] L. Wixson and D. Ballard. Using intermediate objects to improve the efficiency of visual search. *IJCV*, 18(3):209–230, 1994.

[11] Y. Ye and J. K. Tsotsos. Where to look next in 3d object search. In *Proceedings of the IEEE International Symposium on Computer Vision*, Florida, USA, November 1995.