Attention-Path Planning Based on Adaptive Submodular Optimization

Hosun Lee, Sungmoon Jeong, Tokuichi Nakashima, Geunho Lee and Nak Young Chong

***Abstract*— This paper proposes a new attention-path planning algorithm that allows robots with limited sensing coverage to identify an unknown entity efficiently. Our focus is placed on how to plan optimal sequences of views to access more useful information needed to understand the entity. The adaptive submodular optimization technique guaranteed to achieve near-optimal performance is used to maximize the expected information gain. We verified the validity of the proposed approach to the face recognition problem through preliminary experiments.**

I. Introduction

Service robots are expected to support people in their daily life, performing different kinds of tasks adapting to people, situation, and surroundings. Therefore, service robots can be considered as embodied robots with the rudiments of cognition which ensure suitable action/perception planning. For this, they retrieve information from various sources to reach an appropriate decision and action. The major challenge is how to deal with insufficient perception and massive data processing. These limitations may cause wrong decisions that lead to unwanted consequences. To cope with this problem, a new adaptive learning technique is developed to efficiently reduce the uncertainty.

Robots can not usually obtain a whole bunch of information at once. Likewise, it is difficult to process large amounts of information within a certain time frame. Therefore, they need proper action planning to optimally attend to visual information. There have been many researches about the adaptive attention system to understand the human visual system and to improve the computer vision system [1] [2]. Notably, they did not consider the limited sensing coverage.

We deal with the attention path planning problem under the sampling limitation of vision system, where the system obtains only partial information at each sampling step. Then, our goal is how to plan next attentions based on the current information of attention and knowledge to identify the unknown entity. From the fundamentals of information theory, information is defined as the amount of uncertainty reduced by observations. An information gain function modeled to measure the expected information gain is maximized using adaptive submodular optimization framework. Finally, the observation from the current attention is used to build appropriate information to carry out the task. We implement this algorithm to build an unsupervised visual attention-path planning. Sequentially observed information is used to recognize the class of the target face image among the training data set.

The authors are with the School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa, Japan {Hosun_LEE, jeongsm, VirtueMarket, geun-lee, nakyoung}@jaist.ac.jp

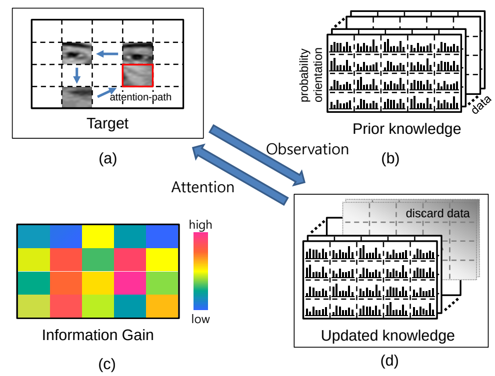


Fig. 1. Framework of attention path planning algorithm

II. Attention-path Planning Algorithm

Algorithm 1: Attention-path planning algorithm

input : prior knowledge $\mathbb{D}$, size of attention grid $[n,m]$
output: set $A \subseteq \mathbb{A}$, set $\mathbb{D}$

begin
  $A \leftarrow \emptyset$, $D_A \leftarrow \emptyset$, $O_A \leftarrow \emptyset$;
  for $i = 1$ **to** $n \times m$ **do**
    foreach $a \in \mathbb{A} \setminus A$ **do**
      compute $\Delta(a|D_A)$;
    end
    Select $a^* \in \arg\max_a \Delta(a|D_A)$;
    Set $A \leftarrow A \cup a^*$;
    Observe $\mathbb{O}(a^*)$;
    Select $D_{dis} :=$ `DiscardData(`O_A,D_A`);`
    Set $\mathbb{D} \leftarrow \mathbb{D} \setminus D_{dis}$;
  end
end

We formulate attention-path planning as an adaptive submodular optimization problem [3], attempting to apply to the face recognition problem as shown in Fig. 1. A brief sketch of the proposed algorithm is given in Algorithm 1.

Initially, a target image consisting of $n \times m$ rectangular patches remain face down and each of the patches will be turned up sequentially as shown in Fig. 1(a). The position of attention is represented with a vector, $[c,r]^T$, where c and r are the column and row positions within a target image. Let $\mathbb{A}$ be the set of all attention positions within a target image, and an attention position $a \in \mathbb{A}$ is selected at each iteration.

As a result of an attention, an observation o, the feature data of the target image, is received from $\mathbb{O}$, the set of all possible observations. Prior knowledge, $\mathbb{D}$, and its feature data, $d \in \mathbb{D}$, also exist as shown in Fig. 1(b). These data are also divided into $n \times m$ rectangular patches. The probability of certain state is represented as $p(d)$. Let D_A and O_A be the selected data and observations when $A \in \mathbb{A}$ is all the attention placed.

We model the information gain function $f(A,D)$ as the dissimilarity variance in a set of training patch images D considering attention positions A. Then, the expected marginal benefit of performing an attention is defined as

$$\Delta(a|D_A) = \mathbb{E}[f(A \cup \{a\}, \mathbb{D}) - f(A, \mathbb{D})|D_A]. \quad (1)$$

Our goal is to find an optimal sequence of attention which maximizes the expected marginal benefit as shown in Fig. 1(c). A simple greedy algorithm is guaranteed to provide near-optimal performance since the information function is modeled to satisfy properties detailed below.

When $X \subseteq Y \subseteq \mathbb{A}$, $a \in \mathbb{A} \setminus Y$,

Adaptive submodularity: The expected marginal benefit of adding an attention to a smaller attention set is at least as much as adding it to the larger attention set.

$$\Delta(a|D_X) \geq \Delta(a|D_Y) \quad (2)$$

Adaptive monotonicity: The expected marginal benefit is nonnegative for all attentions.

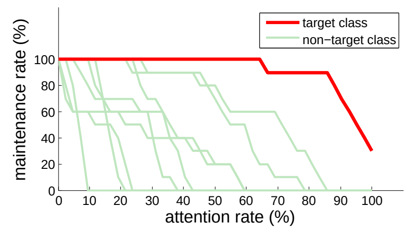
$$\Delta(a|D_X) \geq 0 \quad (3)$$

Achieving an observation from the selected attention, the knowledge is updated by discarding some of data from the current knowledge which have significant dissimilarity from the observation as shown in Fig. 1(d). The system can classify the target image by a simple voting method with the updated knowledge related to the number of remaining data per each class.

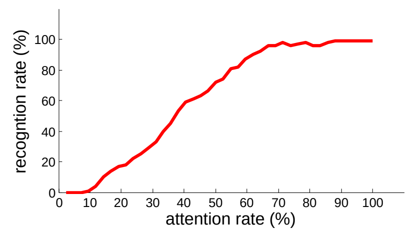
III. Experimental Verification

To demonstrate the effectiveness of the proposed method, the Yale face database B was used, which contains faces under various illumination conditions. Within the database, 10 face classes, each consists of 20 images, are randomly selected. 20 images per each class are randomly separated into two sets: called the training set and the test set. The experimental results were obtained by increasing the percent rate of attention as shown in Figs. 2 and 3.

The performance of face recognition is evaluated by using simple voting between the number of maintenance data in each class from the knowledge model at each time step. In Fig. 2(a), the maintenance rate of non-target class rapidly decreases at each attention by discarding data, yet the maintenance rate of target class is much higher than the others through increasing attention. Fig. 2(b) shows that the recognition performance is improved as the number of attention trials increases. In Fig. 3, the information gain of the early selected attention is higher than that of the late selected attention. The area under the graph means the total amount of information. The system makes information processing less costly by discarding data.



(a)



(b)

Fig. 2. Face recognition according to the attention rate. (a) maintenance rate for a target and non-target classes, (b) average recognition rate with 10 classes.

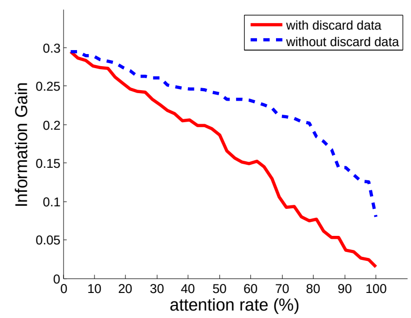


Fig. 3. Information gain with 10 classes according to the attention rate.

IV. Conclusion

A new attention control algorithm was proposed for the face recognition system. The proposed algorithm enabled the system to decide the next attention position using previous observations and updating knowledge to perform the task. Specifically, adaptive submodular optimization was used to maximize the information gain function. We have shown that the proposed model delivers superior classification performance by selecting *informative* attention positions with comparatively low computational cost and resources.

References

[1] L.W. Renninger, P. Verghese, J. Coughlan, "Where to look next? Eye movements reduce local uncertainty," *Journal of Vision*, 7(3):1-17, 2007

[2] Y. Su, S. Shan, X. Chen, W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Transactions on Image Processing*, 18(8):1885-1896, 2009

[3] D. Golovin, A. Krause, "Adaptive submodularity: theory and applications in active learning and stochastic optimization," *Journal of Artificial Intelligence Research*, 42(1):427-486, 2011