

Measuring Task Performance Using Gaze Regions

Irwandi Hipiny

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak.
mhihipni@fit.unimas.my

Hamimah Ujir

uhamimah@fit.unimas.my

Abstract — We present a novel method for measuring task performance using gaze regions, i.e., scene regions fixated by a subject as he or she performs a familiar manual task. The scene regions are learned as a bag of features representation, using library lookup based on the Histogram of Oriented Gradients feature descriptor [1]. By establishing a set of task-specific exemplar models, i.e., models sourced from Pareto optimal sequences, the approach recognizes the local optima within a set of task-specific unlabeled models by estimating the distance (of each unlabeled model) to the exemplar models. During testing, the method is evaluated against a dataset of egocentric sequences, each containing gaze data, belonging to three manual skill-based activities. The results show perfect classification’s accuracy on several proposed schemes.

Keywords-component; Histogram of Oriented Gradients feature descriptor, Gaze Regions, Egocentric

I. INTRODUCTION

The ability to measure task performance has many uses, for example, in a scenario where recognizing the user’s (of an assistive system) expertise level is desired. The recognition allows the assistive system to provide assistance that is compatible to the user’s needs. Another example is auto assignment of greater weights to feature vectors recognized as sourced from well-performed activity sequences during training since they arguably contain less noise (i.e., a truer representation of the activity class).

This paper concerns the recognition of the local optima, i.e., the better-performed sequence, given a set of unlabeled sequences. The sequences were captured using an ASL Mobile Eye device [2], i.e., a dual camera device where an outward-facing RGB camera captures the user’s egocentric view whilst an inward-facing camera captures the movement of the right pupil to track the user’s gaze fixations. The resulting gaze regions are learned as a multi-voting bag of Visual words to improve classification’s accuracy under noisy matching setting and limited training exemplars [3].

Previous literature suggests that task experts (who are more likely to perform a task well) adopt a similar gaze control when performing a task due to the possession of domain knowledge [4–8] and their minimal dependency on visual feedback [9, 10]. Further discussion on these two attributes can be found in Section II. The similar gaze control results to nearly-identical sets of Gaze regions, allowing models arising from the sets to be matched with high correlation. Non-experts are more varied in their gaze control, resulting to a greater distance from the exemplar models learned from the experts. By using several schemes, albeit with a limited dataset, we demonstrate that distance between models built from Gaze regions can reliably be used to measure task performance.

A visual overview of our approach can be found in Fig. 1.

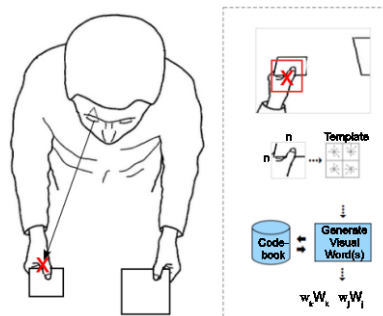


Fig. 1: Gaze regions are learned as a multi-voting bag of Visual words [3]. Multi-voting refers to each template being allowed to match more than one template from the codebook. The matches are weighted and inserted into a bag of features representation, i.e., bag of Visual words. The distance between the built bag and each exemplar bag (built from a task-specific Pareto optimal sequence) is used to estimate how well the task was performed by the subject.

II. RELATED WORK

Previous studies [11–15] in Vision research had found a strong link between gaze and activity. Gaze is found to be involved in the planning and execution of activities in [16], i.e., a person uses gaze to seek out information to serve his or her immediate cognitive goal. Also, gaze fixations fall primarily on task-relevant objects. In [13], gaze was analyzed in a number of simple kitchen-based activities. It was observed that 80% of the gaze fixations fell on task-relevant objects. Another observation made in [17, 18] is that gaze fixations are tightly-linked in time to the evolution of the task.

The use of gaze in computer vision based classification task is rather scarce. It was used for static scene viewing in [19]. In [20, 3], gaze was used to determine the optimal image region for modeling egocentric activities and actions. To the best of our knowledge, our work is the first to use gaze for measuring task performance in egocentric sequences containing gaze data.

A. How Experts Use Gaze

Due to the presence of domain knowledge, experts adopt a similar gaze control when performing a task. According to