Hierarchical Model for Zero -shot Activity Recognition using Wearable Sensors

Abstract

> Unlike conventional models, the proposed framework does not need retraining for recognition of an unseen activity, if the activity can be represented by a combination of the predefined basic actions and objects.
>
> The experimental results showed that the proposed model could recognize three types of activities with precision of 77% and recall rate of 82%, which is comparable to a baseline method based on supervised learning.

Gaze-Guided Object Recognition

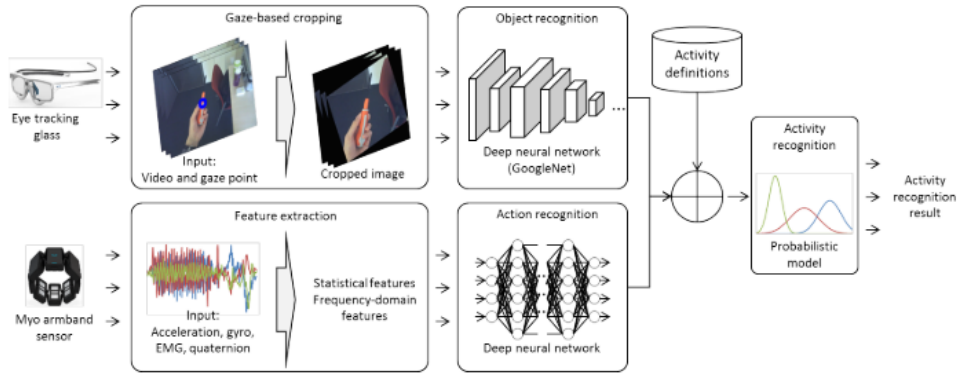


Figure 1: System overview. Object recognition module takes gaze-guided egocentric video and output the probabilities of basic objects. Action recognition module takes multi-modal armband signals and output the probabilities of basic actions. Activity recognition module process these probabilities to output the activity label.

Access visual data by wearable sensors:

- Attach a camera on head/body: 1st person view
- Eye-tracking glasses: gaze point
    - Cropping sub image only around the gaze point

Training Set

- When cropping, randomly change the size of cropping area. Obtain $60 * fps * N_s * N_r$ data in 1 min. (N_s is the different cropping size and N_r is the different degrees of rotation)
- When no objects in the cropped image, define the 'rejected class', which contains all the possible objects and back-ground scene other than the target objects. By this, the model is robust against the False Positive.

Model

- GoogLeNet `Szegedy et al., 2015`
    - **Fine-tuning the last two layers** with above training data.

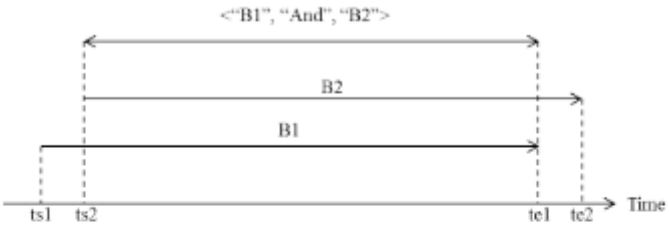
Action Recognition

Training Set

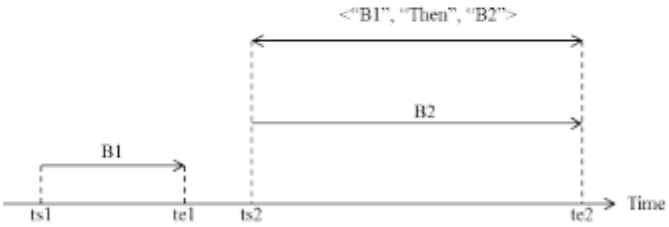
- All data collected by *Myo armband*: quaternion, acceleration, gyro and EMG data are utilized.
    - Statistical features such as maximum, minimum, mean and standard deviation as well as the features in frequency domain, namely, amplitude spectrum obtained by applying fast Fourier transform (FFT) are used.
- The “reject class” is defined in the action recognition model as well to deal with the case when no target action is performed.

Activity Recognition

Strategy



(a) "And"



(b) "Then"

Figure 3: Periods that the conjunction word “And” and “Then” represent.

- Define activities using name of object, name of action, and the conjunction words.

Model

- a probabilistic framework
    - The activity recognition module receives the array o f probabilities from basis recognition module, each of which represents the likelihood of each target object or action.

$$P(activity|s; def(activity)) = p(activity|obj, act, def(activity)) * p(obj|s) * p(act|s)$$

Evaluation

Experiment

Data collection procedure

1. Start recording data.
2. A subject performs one activity 3 to 5 times in a row with short interval between each performance.
3. Stop recording data.
4. Restart recording data.
5. The subject performs the 2nd activity 3 to 5 times in a row.
6. Stop recording data.
7. Iterate the same procedure for the last activity.
8. Iterate the same procedure for the other subjects.、

Table 1: Evaluation data for activity recognition method.

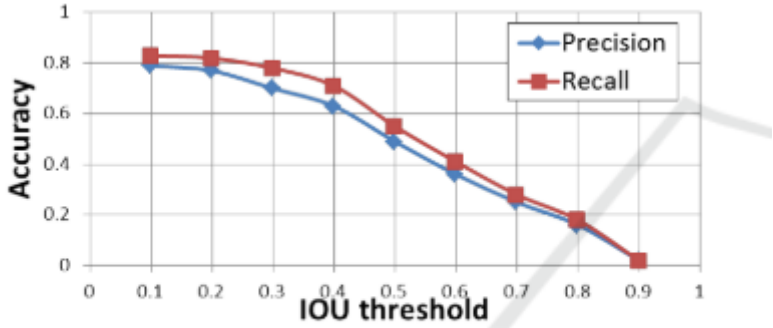
Number of subject	12
Number of target classes	4
Target classes	Putting a bag on a table, Opening a lid of a bottle, Tightening a screw, Others(reject class)
Number of data	Total: 131 Putting a bag on a table: 39, Opening a lid of a bottle: 50, Tightening a screw: 42

Table 2: Definition of the activities.

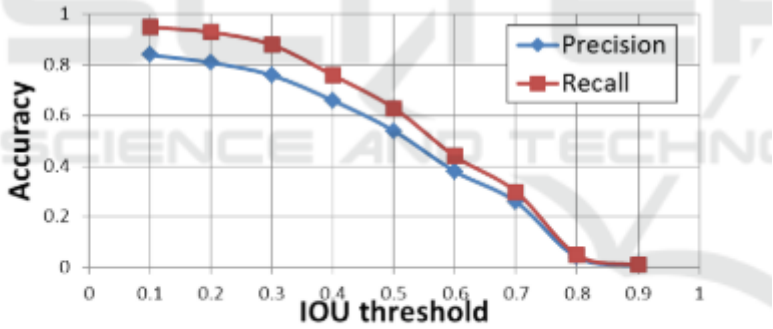
Putting a bag on a table	<“Bag”, “Then”, “Holding”>
Opening a lid of a bottle	<“Bottle”, “Then”, “Twisting”>
Tightening a screw	<“Screw driver”, “And”, “Twisting”>

Result

- Each estimation was regarded as right if the IOU is more than a threshold.
    - Baseline case is the normal supervised learning of SVM.



(a) Proposed (zero-shot)



(b) Baseline (many-shots)

Figure 5: Precision and recall rate for different IOU thresholds.

Table 7: Evaluation result of the activity recognition method. Threshold for IOU: 0.2.

Activity	Precision	Recall
Putting bag	0.96 (27 / 28)	0.69 (27 / 39)
Opening lid	0.72 (44 / 61)	0.88 (44 / 50)
Tightening screw	0.73 (43 / 59)	0.88 (37 / 42)
Total	0.77 (114 / 148)	0.82 (108 / 131)