

Situation awareness via sensor-equipped eyeglasses

Jens Windau, Laurent Itti

Abstract—New smartphone technologies are emerging which combine head-mounted displays (HMD) with standard functions such as receiving phone calls, emails, and helping with navigation. This opens new opportunities to explore cyber robotics algorithms (robotics sensors and human motor plant). To make these devices more adaptive to the environmental conditions, user behavior, and user preferences, it is important to allow the sensor-equipped devices to efficiently adapt and respond to user activities (e.g., disable incoming phone calls in an elevator, activate video recording while car driving). This paper hence presents a situation awareness system (SAS) for head-mounted smartphones. After collecting data from inertial sensors (accelerometers, gyroscopes), and video data (camera), SAS performs activity classification in three steps. Step 1 transforms inertial sensor data into a head orientation-independent and stable normalized coordinate system. Step 2 extracts critical features (statistical, physical, GIST). Step 3 classifies activities (Naive Bayes classifier), distinguishes between environments (Support Vector Machine), and finally combines both results (Hidden Markov Model) for further improvement. SAS has been implemented on a sensor-equipped eyeglasses prototype and achieved high accuracy (81.5%) when distinguishing between 20 real-world activities.

I. INTRODUCTION

Head mounted displays (HMDs) embedded in eyeglasses are the next innovation along the path of communication techniques. Such devices are hand-free systems. Although this is not a new idea, currently released and commercially available products (such as the Project Glass by Google) show the immense potential of this technology. They function as stand-alone computers; their light glass frame is equipped with a variety of sensors; a projector displays images and information onto the eye. While wearing these eyeglasses, the user is continuously exposed to the displayed information. To avoid bombarding a user with unwanted clutter, a management system is required. This management system identifies important messages (e.g., phone calls, email popups) that are relevant to a user's situation. To evaluate the importance of a message, however, knowledge of the user's activity is essential. Thus the challenge of this paper is to present an effective approach to classify activities from sensor data of the eyeglass frame. The data is recorded by a sensor cluster of IMU sensors (accelerometers, gyroscopes) and a camera. Beyond wearable smartphones, the proposed algorithms have potential applications for other types of embedded and robotics systems (e.g., sensors mounted on a robotic vehicle operating over different kinds of terrain).

Although, the functionality of eyeglasses is similar to that of a smartphone, the interaction is often only limited to voice

commands. Instead of, e.g., rejecting messages each time via voice commands, the management system could automatically take over decision control in certain daily situations [Fig. 1]. Adapting these settings and filtering information will guarantee a high usability of the eyeglasses. Furthermore, logging classified activities throughout the day allows support for many other fields. One example is delivering activity information to health care programs for obesity prevention, treatment, or healthy life style recommendations [1].

Eyeglasses Settings	Predicted class																			
	biking	brushing teeth	computer working	doing laundry	driving car	eating	folding laundry	jogging	lying down	playing cello	playing piano	playing table tennis	reading book	running	taking elevator	taking stairs	vacuum cleaning	walking	washing dishes	wiping cupboard
Sample Actions																				
silent volume																				
normal volume																				
high volume																				
no incoming calls																				
no email popups																				
posting location																				
recording video																				

Figure 1. Sample decisions of an individual profile of an eyeglass user

Studies on activity classification have been demonstrated with a variety of sensors, mostly inertial sensors (accelerometer, gyroscope), but also in combination with pressure sensors [1] or a camera [9]. This paper also presents a multi-sensor approach that consists of an accelerometer, a gyroscope and a camera.

Research on activity classification has been done by placing inertial sensors on various body parts: thigh, hip, waist, forearm, chest, lower back, knee, ankle, neck, or foot [1], but rarely head [4][5]. Most work claims that placements around the waist are the closest to the center of mass and thus better represent human motion and activities [5]. Most studies for this scenario were performed with small databases ranging from five to nine activities. Five activities were successfully recognized by a wrist-worn accelerometer (94.13 %) [7] and a waist-worn accelerometer (99.5%) [6]. Six activities achieved also good results with a pocket-worn (91.7 %) [10] and a belt-worn (82.8 %) [2] smartphone accelerometer. An IMU placed at the front hip was able to differentiate between nine activities (90%) [8]. Activities from large databases were performed with distributed inertial sensor networks (up to six sensor modules) with sensor data simultaneously collected from multiple body parts e.g. hip, wrist, arm, ankle, and thigh [5]. The most promising distributed sensor networks scored with almost 84% accuracy on 20 different activities [4]. An inertial sensor network combined with a camera succeeded with 61% on 29 kitchen activities [9]. For fewer activities (up to six) most of the sensor networks achieved results in the 85-95% range [4].

Jens Windau, Laurent Itti are with the iLab and Computer Science Department at the University of Southern California, Los Angeles CA 90089-2520, USA. (website: <http://ilab.usc.edu> phone: +1(213)740-3527 email: jenswindau@gmail.com, itti@pollux.usc.edu)

Head placements are rare for activity classification, and have only been used for detecting falls and measuring balance during walking [5]. One possible reason is that sensor data is strongly affected by variations of head orientation. However, when effectively compensated for head movements, sensor data from eyeglasses have one significant advantage over other body placements: People are wearing eyeglasses with only slight position variations. Smartphones, however, can be put into the right or left pocket, in a purse or backpack. For daily use, sensor-equipped clothes or shoes are less practical because they are often changed. Eyeglasses are convenient particularly for people wearing prescription eyewear anyway. Furthermore, eyeglasses are self-contained and do not need a wired or wireless connection to sensors distributed all over the body, such as sensor networks do. The approach presented in this paper is also an eyeglass solution, but it has a set of unique contributions: (1) a face-mounted design of a sensor cluster to record sensor data, (2) an effective compensation technique for head movements, (3) the identification of critical features, (4) a large database with previously identified activities, (5) a classification technique to use the above ideas to classify and learn future activities in a short time.

II. THE SAS APPROACH

(1) Placement and design of the sensor cluster

The head placement has certain benefits when recording data with a sensor cluster. From a view point of inertial sensors, the head captures full-body motions (e.g. walking, jogging) with similar results as other body parts. Part-body motions (e.g. washing dishes, eating) can only be captured indirectly. When measuring part-body motions, the sensors actually record small counter-motions of the head. These counter-motions appear when the upper body compensates for lower body movements (e.g. lifting a hand). This is the reason why also part-body motions can be observed with inertial sensors. From the view point of a camera, the head is an optimal location to collect activity data. The camera is mostly directed to where the eyes pay attention to and do not have to deal with occlusion problems (such as smartphones in backpack, pocket, or purse). From an ergonomic point of view, the eyeglasses allow a very comfortable and easy daily use. Both IMU sensors and camera are integrated in the left earpiece of our prototype IMU+Camera eyeglasses [Fig 2].



Figure 2. Design of the eyeglasses. The sensors are attached to the left earpiece

The design of the eyeglasses is reduced to two minimum necessary sensors to classify an activity: A Pololu MinIMU-9 v2 Inertial Measurement Unit captures activities via accelerometer and gyroscope. A 160x120 pixels camera records images of the environment to put classified activities into the right contextual setting (e.g. indoors, outdoors).

(2) IMU data transformation into stable coordinate system

One contribution of SAS is to extract features that are not affected by the orientation of the head or the exact placement of the glasses on the head. When a person performs a head movement, the local sensor coordinate system will change accordingly. The key idea of the SAS approach is to keep the sensor data in a coordinate system that will stay stable even when the head moves. Thus the sensor data needs to be transformed from its dynamic local sensor coordinate system into a normalized coordinate system. This normalized coordinate system is defined as the x-axis pointing out of the eye, the y-axis points out of the right ear and the z-axis points vertically down. The normalized coordinate system compensates for any head orientation variation in roll and pitch angle, but ignores the yaw angle. The reason for ignoring the yaw angle is that the direction in which an activity is performed is not relevant. For example, walking north-bound (yaw = 0°) should result in the same as walking south-bound (yaw = 180°). When variations in roll and pitch angle happen, the accelerometer data will be significantly affected by gravitational acceleration. One can walk with the head facing up, down, or tilted to the side [Fig. 3].

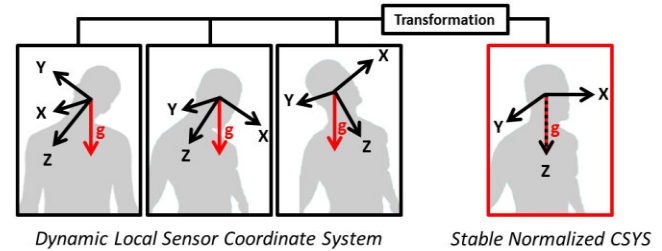


Figure 3. Sensor data from any head orientation is transformed in a stable normalized coordinate system

The local sensor coordinate system will change its orientation along with the head orientation. Particularly accelerometer data will be influenced by different head orientations [Fig 4].

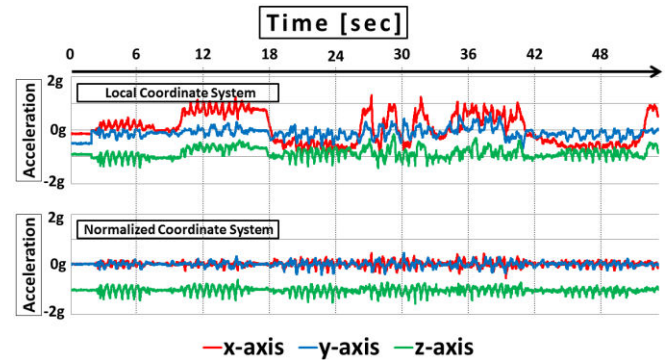


Figure 4. Sequences of walking with different head orientations.

As a result, for example, statistical or physical features, such as mean, variance, or energy will show very different results for each head orientation. Most likely it will also influence the performance of the classifier. The transformation into a stable normalized coordinate system will eliminate this effect and will compensate for any head orientation changes. The physical idea behind the transformation is to use the gravity vector (g) as a static reference. This vector always points vertically down and aligns with the z-axis of the normalized coordinate system. The g -vector can be measured via the accelerometer. Furthermore, gyroscope input and fusion techniques (DCM Algorithm) are used to make the calculation of the g -vector very robust – even when measuring fast motion activities (e.g. jogging, running). When transforming, the z-axis of the local sensor coordinate system is rotated until aligned with the g -vector. In other words, it is now aligned with and transformed into the normalized coordinate system. As a result, the roll and pitch angle of the new coordinate system will be 0° at all times independent from the head orientation. Now the sensor data is completely invariant from the orientation of the head and prepared for feature extraction.

(3) Identification of critical features

Feature extraction is separately performed for inertial sensor data and video data.

When extracting features from inertial sensor data, SAS differentiates between statistical and physical features. Statistical features are purely statistical operations (e.g. mean, variance) on sensor data. Physical features, however, are calculated from an equation with physical meaning (e.g. movement intensity, energy consumption). Previous research work uses either pure statistical features [9] or both statistical and physical features combined [8]. Feature extraction from video data is limited to one feature per image. Each feature is a large GIST vector that describes the context of a scene.

(4) Creating an extendable database

SAS must be trained with activities and types of environments in advance to classify them. First, one minute of inertial sensor data was recorded for each of the 20 activities. 22-dimensional feature vectors were extracted from six second windows and stored in the activity database. Activities that are composed of many different motions (e.g. folding clothes) tend to widely spread out. Activities with one clearly defined motion (e.g. brushing teeth) mostly accumulate in small local areas [Fig.5]. In addition, SAS can

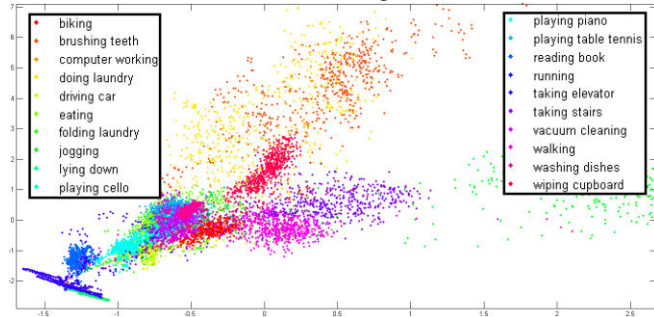


Figure 5. Dimension-reduced activity features displayed in 2D via principal component analysis. Each color represents a different activity

easily extend the activity database. Each new activity requires only one minute of inertial sensor data, to calculate sufficient new features for classification.

Second, 3000 images of video data of two types of environments were recorded (2100 indoors, 900 outdoors). A 512-dimensional GIST vector was extracted from each image and stored in the environment database [Fig. 6].

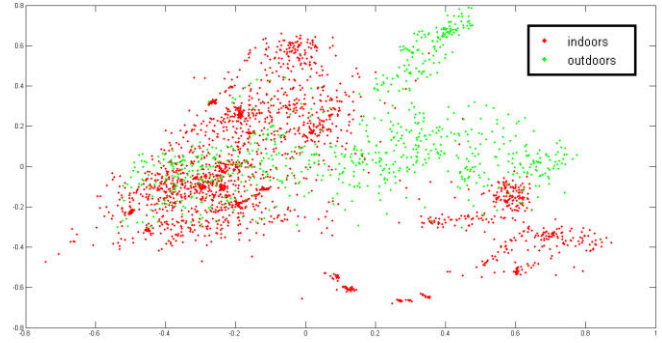


Figure 6. Dimension-reduced environment database in 2D via principal component analysis with two classes (inside, outside)

(5) Classification technique for activities

First, a Naive Bayes classifier was trained with the activity database. It compares incoming activity features with database features and outputs a list of probabilities of matching activities. Second, a Support-Vector-Machine is trained with the environment database. It compares incoming environment features with database features and outputs the matching type of environment.

III. THE MEASUREMENT IN DETAIL

A. Overview

SAS records IMU data with a frequency of 50 Hz, camera images with 2.5 Hz. Furthermore, the architecture can be structured in three major processing steps [Fig. 7]. Step 1 prefilters sensor data and transforms it from a local dynamic coordinate system into the stable normalized coordinate system. Step 2 handles the feature extraction for both IMU and camera data. IMU data is segmented into windows, followed by the extraction of statistical and physical features from each window. For each image in the camera data, one GIST feature is calculated. Step 3 is performing classification (activities, environments) by using a network of multiple classifiers. The result of SAS is a list of activities with assigned probabilities.

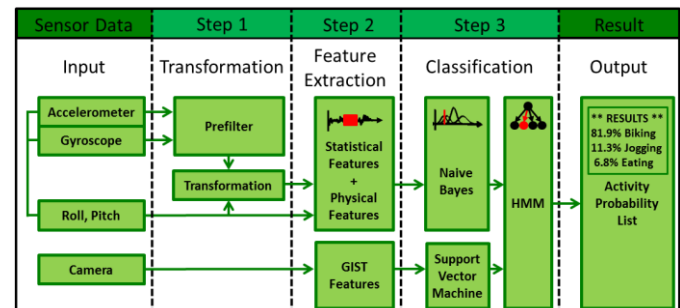


Figure 7. Overview of the system architecture

B. Steps in detail

Transformation (only for IMU data)

In the transformation step, the raw sensor data is first pre-filtered (error-corrected, offset compensated, and resampled to a constant 100 Hz). Then, the data is transformed into the normalized coordinate system. The transformation is accomplished with the following two equations:

$$\begin{aligned} \mathbf{a}_{normalized} &= \mathbf{R}_a^{-1} \cdot \mathbf{a}_{sensor} \\ \omega_{normalized} &= \mathbf{R}_\omega^{-1} \cdot \omega_{sensor} \end{aligned}$$

When recorded in the local coordinate system, acceleration data is stored in the 3x1 vector $\mathbf{a}_{sensor}$, with a_x , a_y , a_z representing the acceleration along the sensor axes. Gyroscope data is stored in the 3x1 vector ω_{sensor} , with ω_x , ω_y , ω_z defined as the angular velocities along the sensor axis.

$$\mathbf{a}_{sensor} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \quad \omega_{sensor} = \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

In the normalized coordinate system, transformed acceleration data is stored in $\mathbf{a}_{normalized}$ with a_x , a_y , a_z representing the acceleration along the normalized coordinate axes. Gyroscope data is stored in $\omega_{normalized}$ with ω_x , ω_y , ω_z and is equal to the first derivative of the Euler Angles roll ϕ , pitch θ , and yaw ψ .

$$\mathbf{a}_{normalized} = \begin{pmatrix} a_X \\ a_Y \\ a_Z \end{pmatrix} \quad \omega_{normalized} = \begin{pmatrix} \omega_X \\ \omega_Y \\ \omega_Z \end{pmatrix} = \begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix}$$

In order to connect the sensor data vectors ($\mathbf{a}_{sensor}$, $\mathbf{a}_{normalized}$) and (ω_{sensor} , $\omega_{normalized}$), two transformation matrices $\mathbf{R}_a$ and $\mathbf{R}_\omega$ are required. Both are composed from basic rotation matrices:

$$\mathbf{R}_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & \sin(\phi) \\ 0 & -\sin(\phi) & \cos(\phi) \end{bmatrix} \quad \mathbf{R}_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix}$$

The transformation matrix $\mathbf{R}_a$ for transforming accelerometer data into normalized coordinates is derived as following:

$$\mathbf{R}_a = \mathbf{R}_x(\phi)\mathbf{R}_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ \sin(\theta)\sin(\phi) & \cos(\phi) & \cos(\theta)\sin(\phi) \\ \sin(\theta)\cos(\phi) & -\sin(\phi) & \cos(\theta)\cos(\phi) \end{bmatrix}$$

The transformation matrix $\mathbf{R}_\omega$ to transform gyroscope data into normalized coordinates is derived as following [13]:

$$\begin{aligned} \mathbf{R}_\omega &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{R}_x^T(\phi) \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{R}_x^T(\phi) \cdot \mathbf{R}_y^T(\theta) \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & -\sin(\theta) \\ 0 & \cos(\phi) & \sin(\phi)\cos(\theta) \\ 0 & -\sin(\phi) & \cos(\phi)\cos(\theta) \end{bmatrix} \end{aligned}$$

Once the transformation into the stable normalized coordinate system is performed, features are extracted from sensor data.

Feature Extraction for IMU data

In total, 22 features are extracted from inertial sensor data [Fig. 8]. Seven features (energy, periodicity) are created by transforming sensor data into the frequency spectrum via FFT (Fast Fourier Transformation).

Statistical Features		
Type	Sensor Data	Dimensions
Mean	Accelerometer	3
	Gyroscope	3
Variance	Accelerometer	3
	Gyroscope	3
Total Number of Statistical Features		12
Physical Features		
Type	Sensor Data	Dimensions
Movement Intensity	Accelerometer	2
Energy	Accelerometer	4
Energy Consumption	Accelerometer	1
Periodicity	Gyroscope	3
Total Number of Physical Features		10
Total Number of Features		22

Figure 8. Statistical and Physical Feature distribution

Mean is used to measure the average acceleration and angular velocity for each sensor axis over one window length. Mean is larger for activities with strong body motions. **Variance** describes how far acceleration and angular velocity are spread out along an axis. Fast and wide motions are larger. **Movement Intensity** (MI) specifies the

$$MI[t] = \sqrt{a_x[t]^2 + a_y[t]^2 + a_z[t]^2}$$

intensity of motions. Mean and variance of MI are calculated over one window [8]. **Energy** describes the motion quantity [8]. E_i measures the energy for each axis; E is the energy over the entire system. Parameters are N (number of samples

$$E_i = \frac{1}{N} \sum_{f=0}^F M_{ai}[f]^2 \quad E = \frac{1}{N} \sum_{f=0}^F (M_{ax}[f]^2 + M_{ay}[f]^2 + M_{az}[f]^2)$$

per window length), M_{ai} (discrete FFT component magnitude of acceleration along the axis i), f (frequency), and F (maximal frequency of window). **Energy Expenditure** (EE) is also known as the normalized signal

$$EE = \frac{1}{T} \left(\sum_{t=1}^T |a_X[t]| + \sum_{t=1}^T |a_Y[t]| + \sum_{t=1}^T |a_Z[t]| \right)$$

magnitude area [8] and describes the amount of energy used for an activity. T is specified as the time of one window.

Periodicity (f_{peak}) detects recurring motions. f_{peak} determines

$$f_{peak} = \text{MAX}(M_{\omega_i} > c)$$

the highest dominant frequency for an axis i . Parameters are M_{ω_i} (discrete FFT component magnitude of angular velocity along the axis i), c (minimum required magnitude threshold) to avoid noise peaks.

Feature Extraction for video data

Camera data is processed with a GIST descriptor for scene recognition provided by A. Torralba from MIT [11]. For each image, one 512 dimensional GIST feature is calculated.

Classification

SAS uses the extracted features to classify a situation. A situation is specified by an activity and an environment. For this reason, the classification process is split into three parts. First, a Naive Bayes classifier distinguishes activities by analyzing inertial data. In parallel, a Support Vector Machine (SVM) classifies between environments via video data. Finally, both results are combined in a Hidden Markov model (HMM) to further improve the activity classification.

The Naive Bayes Classifier (activity classification) is trained with one minute of training data per activity. This training data is divided into 10 training windows – each 6 seconds. A Naive Bayes Classifier only requires a relatively small amount of training data to estimate its two internal parameters (mean, variance) for classification. The **SVM** (environment classification) is an effective classifier for high dimensional data [8][12]. SVM is trained with GIST features of 1680 images (1260 indoors and 420 outdoors). The **HMM** combines both classification results (activities, environment) and used as observation inputs in the HMM. The number of possible observations (40) is the amount of activities (20) times the amount of environments (2) [Fig. 9]. Furthermore, HMM is used to model transition probabilities between activities. Therefore, the matrices with transition and emission probabilities were designed based on real-world behavior: Activity states are more likely to repeat themselves than switching into a different activity state. Indoor activities are more likely to transfer into indoor activities. This applies to outdoor activities accordingly. Overall, the walking state is a more likely transfer state, because it connects most activities in the real world.

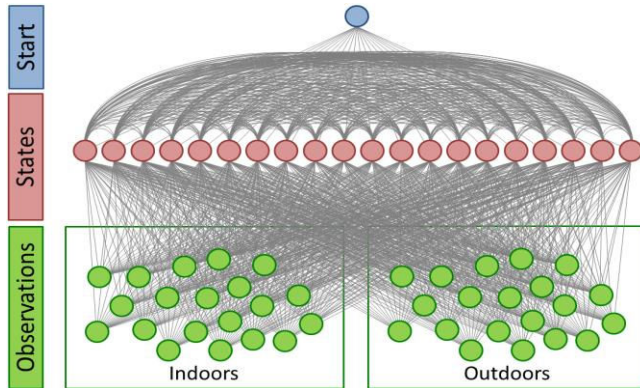


Figure 9. HMM architecture (40 observation states, 20 hidden states)

IV. TEST RESULTS

A. Performance tests of Naive Bayes and SVM

SAS was tested with data chunked into 6s windows with 3s overlap. Each window was classified. The training and testing was performed on different days by the same subject. Test data only included activities which the database was previously trained for [Fig. 10]. The device may need to be trained for each further subject in future (similar to speech recognition). First, the classification performance for activities (Naive Bayes Classifier) was tested with sets of 5, 10, 15 and 20 activities. Then, the success rate of the SVM classifier for environments was measured. Finally, the fusion performance of the HMM was determined.

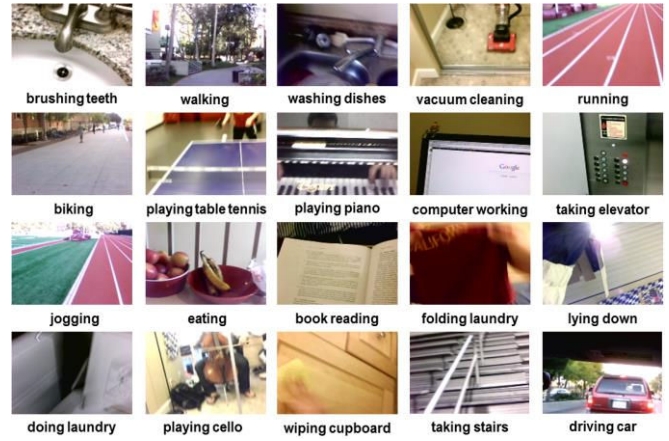


Figure 10. The 20 daily activities used for testing

Activity classification with inertial data: The performance of SAS was first tested with a small set of 5 basic motion activities (**lying down, walking, jogging, biking, running**). Then, 10 activities were tested by adding mainly indoor tasks (**vacuum cleaning, washing dishes, brushing teeth, taking stairs, eating**) followed by 15 activities by including hobby tasks (**playing cello, playing piano, playing table tennis, computer work, reading book**). Finally, all 20 activities were tested by adding complex activities (**folding laundry, doing laundry, wiping cupboard**) or activities (**driving car, taking elevator**) that had a very similar head motion pattern to an activity already existing in the database (computer work).

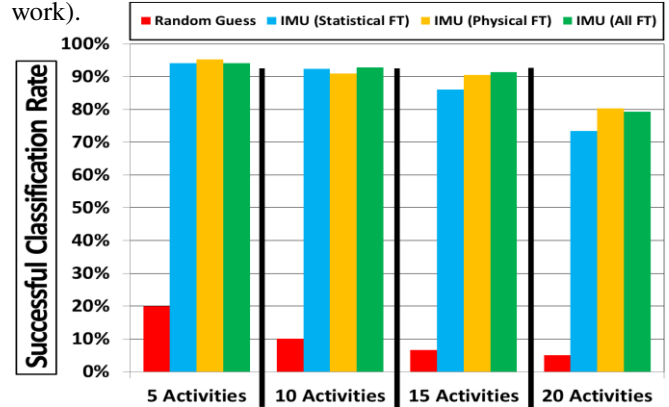


Figure 11. Test results of activity sets with pure inertial data

Results show that pure physical features and all features combined show a slightly stronger performance than pure statistical features. For large databases (20 activities), pure physical features (80.3%) even showed a slightly better performance than all features combined (79.3%) [Fig. 11]: It turns out that statistical features do not improve the overall result of physical features. Therefore, only physical features were used for further testing with HMM.

Environment classification with video data: The SVM classifier over GIST data showed a successful performance of 89.8% correct classification.

B. Full system testing (including HMM)

The overall classification performance of SAS was tested. For this reason, both outputs of Naive Bayes (activities) and SVM (environments) were assigned to the HMM. The HMM re-classified activities to take environment information into account. The HMM classification performance is shown in a confusion matrix (in %) [Fig. 12].

Confusion Matrix [%]	Predicted class																			
	biking	brushing teeth	computer working	doing laundry	driving car	eating	folding laundry	jogging	lying down	playing cello	playing piano	playing table tennis	reading book	running	taking elevator	taking stairs	vacuum cleaning	walking	washing dishes	wiping cupboard
biking	90	0	0	0	0	0	2	0	0	0	0	0	6	0	2	0	0	0	0	0
brushing teeth	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
computer working	0	0	79	0	4	2	0	0	12	0	0	0	0	0	3	0	0	0	0	0
doing laundry	0	0	0	98	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
driving car	6	0	14	0	34	0	20	0	0	0	0	0	0	0	26	0	0	0	0	0
eating	0	2	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	2	0
folding laundry	0	0	0	0	4	0	86	0	1	7	0	0	0	0	2	0	0	0	0	0
jogging	0	0	0	0	0	0	0	97	0	0	0	0	0	0	2	0	1	0	0	0
lying down	0	0	0	0	0	0	8	0	91	0	0	0	0	0	1	0	0	0	0	0
playing cello	0	3	0	7	0	0	8	0	0	82	0	0	0	0	0	0	0	0	0	0
playing piano	0	0	0	0	0	0	5	0	0	0	95	0	0	0	0	0	0	0	0	0
playing table tennis	1	0	0	16	0	0	2	0	0	0	79	0	0	0	2	0	0	0	0	0
reading book	0	0	1	0	0	0	0	1	0	0	0	91	0	7	0	0	0	0	0	0
running	0	0	0	0	0	0	3	0	0	0	0	0	96	0	1	0	0	0	0	0
taking elevator	0	4	0	0	0	5	9	0	5	11	0	0	0	66	0	0	0	0	0	0
taking stairs	0	0	0	0	0	0	8	0	0	0	0	0	0	0	92	0	0	0	0	0
vacuum cleaning	0	0	0	0	0	0	7	0	0	0	0	0	0	2	0	91	0	0	0	0
walking	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	85	0	0	0	0
washing dishes	0	9	0	0	0	1	8	0	0	0	1	0	0	0	0	0	81	0	0	0
wiping cupboard	52	0	0	1	0	0	5	0	0	0	0	1	0	0	37	0	0	0	0	4

Figure 12. Hidden Markov Model Results: Correct classification (green), wrong classification (red), transition errors (orange), initializing error (blue)

HMM achieved a classification result of 81.5%. Because of the transition probabilities and the SVM input, the activities were recognized in a significantly more stable way than only using Naive Bayes. This greater temporal stability was the main added value of the HMM. However, the classification still fails in certain situations. For example, stopping a car at an intersection will naturally discontinue the characteristic sensor data pattern for “driving car”. When the stop lasts for too long, the classified activity is likely to switch from “driving car” into a better matching sensor data pattern, e.g. “computer work” or “taking elevator” (during constant-speed travel, and discounting the initial / final vertical acceleration phases) [Fig. 12]. Under certain circumstances, classifying an activity can completely fail, if an activity can be performed in too many different ways. While “wiping cupboard” in certain directions, head counter motions apparently showed a too high similarity to “biking” and “taking stairs”. As a result, the HMM misclassified it. Furthermore, unstable and frequent switching of the Naive Bayes classifier and low HMM transition probabilities were further contributing to the misclassified state [Fig. 12]. However, SAS shows a very stable and promising overall result with most activities being classified with 85% and higher success rate. Interestingly, this result was obtained when using the physical features only, as we found that the contribution from statistical features was negligible. A sample video was recorded for all 20 activities and demonstrates the classification outcomes of all three classifiers [see attached video].

V. FUTURE IMPROVEMENTS AND CONCLUSION

Activities without a distinct motion (computer work, taking elevator), or with too wide a range of possible submotions (wiping cupboard, folding laundry), are easily misclassified. Besides extracting physical features, the composition of these activities needs to be further studied. If a complex activity such as folding clothes can be broken into all submotions, a probability distribution (e.g. 40% shaking, 20% vertical movement, 40% horizontal movement) could possibly lead to

a more promising result. When extending the database with more activities, the video data need to further be used to distinguish between significantly more environments or even object recognition. Often, eyeglasses come along with microphones. Additional sensor input could also contribute by measuring the surrounding noise level (e.g. car noises in street, sound of music instruments, pure silence).

SAS showed that head-mounted sensor systems can be successfully used to measure 20 activities with an accuracy of 81.5% (15 activities with 91.3%). With the new eyeglass devices entering the market of communication technology, situation-adapted controlling of system settings guarantees a high-usability for daily use. By providing valuable activity classification, we hope that SAS made a contribution to the development of situation awareness systems for eyeglasses.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation (grant number CMMI-1235539), the Defense Advanced Research Projects Agency (N65236-12-C-3884), the Army Research Office (W911NF-11-1-0046 and W911NF-12-1-0433), U.S. Army (W81XWH-10-2-0076), and Google. The authors affirm that the views expressed herein are solely their own, and do not represent the views of the United States government or any agency thereof.

REFERENCES

- [1] T. Zhang, W. Tang, “Classification of Posture and Activities by Using Decision Trees”, 34th Annual International Conference of the IEEE EMBS, San Diego, CA, USA, 2012
- [2] S. Zhang, P. McCullagh, C. Nugent, H. Zheng, “Activity Monitoring Using a Smart Phone’s Accel. with Hierarchical Classification”, 6th Intl. Conf. on Intellig. Environments, Kuala Lumpur, Malaysia, 2010
- [3] X. Long, B. Yin, R.M. Aarts, “Single-acc.-based daily physical activity classif.”, 31st Annual Int. Conf. of the IEEE Eng. in Med. and Bio. Society (EMBC 2009) 6107-6110 Minneapolis, MN, USA, 2009
- [4] L. Bao, S. Intille, “Activity Recognition from User-Annotated Acceleration Data”, PERSASIVE 2004, vol. LNCS 3001, A. Ferscha, F. Matterns, Eds. Springer-Verlag Berlin Heidelberg, 2004, pp. 1-17
- [5] C. Yang, Y. Hsu, “A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring”, Sensors vol. 10, 2010, pp. 7772-7788
- [6] M. Lee, J. Kim, K. Kim, I. Lee, S.H. Jee, S.K. Yoo, “Physical Activity Recognition Using a Single Tri-Axis Accelerometer”, World Congress on Engineering and Comp. Sc. 2009 Vol I, San Francisco, USA, 2009
- [7] S. Chernbumroong, A.S. Atkins, H. Yu, “Activity classification using a single wrist-worn accelerometer”, 5th IEEE Intl. Conf. on Software, Knowledge, Information Mgmt and App., Benvento, Italy, 2011
- [8] M. Zhang, A.A. Sawchuk, “A Feature Selection-Based Framework for Human Activity Recognition Using Wearable Multimodal Sensors,” Intl. Conference on Body Area Networks, Beijing, China, 2011
- [9] E.H. Spriggs, F. Torre, M. Hebert, “Temporal Segmentation and Activity Classification from First-person Sensing”, IEEE Workshop on Egocentric Vision, Miami, FL, 2009, pp. 17-24
- [10] J.R. Kwapisz, G.M. Weiss, S.A. Moore, “Activity Recognition using Cell Phone Accelerometers”, 4th Intl. Workshop on Knowledge Discovery from Sensor Data, Washington, DC, USA, 2010
- [11] A. Oliva, A. Torralba, “Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope”, Intl. Journal of Comp. Vision 42(3), Kluwer Academic Publishers. Netherlands, 2001, pp. 145-175
- [12] C-W. Hsu, C.C. Chang, C.J. Lin, “A Practical Guide to Support Vector Classification”, Technical Report, Computer Science Department, National Taiwan University, 2003
- [13] M. Cook, “Flight Dynamics Principles: A Linear Systems Approach to Aircraft Stability”, 2nd Edition, 2007, Butterworth-Heinemann, Chapter: “Angular velocities transformation”, p. 23