User Behaviour Recognition for Interacting with an Artistic Mobile Application

Jody Hausmann*, Kevin Salvi, Jérôme Van Zaen, Adrian Hindle and Michel Deriaz
Institute of Services Science
Faculty of Economic and Social Sciences
University of Geneva, Switzerland
{jody.hausmann, kevin.salvi, jerome.vanzaen, adrian.hindle, michel.deriaz}@unige.ch

***Abstract*—Interacting with smartphones generally requires direct input from the user. We investigated a novel way based on the user's behaviour to interact directly with a phone. In this paper, we present MoveYourStory, a mobile application that generates a movie composed of small video clips selected according to the user's position and his current behaviour when the user is moving. Towards this end, we have implemented an activity recognition module that is able to recognise current activities, like walking, bicycling or travelling in a vehicle using the accelerometer and the GPS embedded in a smartphone. Moreover, we added different walking intensity levels to the recognition algorithm, as well as the possibility of using the application in any position. A user study was done to validate our algorithm. Overall, we achieved 96.7% recognition accuracy for walking activities and 87.5% for the bicycling activity.**

Keywords—Context-awarness, Activity Recognition, GPS, Accelerometer, Smartphone

I. INTRODUCTION

Understanding and detecting physical patterns using smartphones is widely used by eHealth software. However, this kind of detection is mainly passive and does not interact with the software itself.

Most of the recent related research on activity recognition is using sensor-embedded smartphones. However, only a few allow multiple device positions and orientations. For example, Stewart et al. [1] used a classifier to detect two levels of walking and running from the accelerometer signal. The device was fixed on the user's upper arm. In [2], the authors experimented several device positions in a pocket for assessing current activities using a classifier based on the accelerometer and the orientation sensor. They detected several activities like walking, bicycling, driving and ascending/descending stairs. The same activities were detected by the approach presented in [3] using a new fuzzy model when the device is carried in a pocket.

The combination of an activity recognition system with a social medias was experimented in [4]. They designed an application called *CenseMe* that combines "sensing presence" (e.g., walking and sitting) with social networks. A similar approach called *crowdsensing* is also described in [5], where a framework called MOSDEN for capturing and sharing sensed data is presented.

Some research uses GPS to monitor activities in combination with the accelerometer. The GPS can be used for detecting transportation modes and thus the current activity. However, GPS is not available indoor, and the signal may be lost in some cases, such as the proximity of buildings. Moreover, the GPS provides noisy data at low speed (e.g., when the user is walking). Finally, with the exception of the speed feature, it is not possible to do a distinction among transportation modes. Hemminki et al. [6] presented a new approach to detect transportation modes using the accelerometer signal only. In [7], they experimented different approaches using accelerometer, GPS or a combination of both. The paper concluded that the combination of both resulted in the best performance.

All of these research articles reported high accuracies for detecting activities, but most of them required a specific or a fixed device position.

In this paper, we propose a novel approach to interact with a smartphone application, using the movements and the behaviour of the user. We developed an application called MoveYourStory (MYS) that uses the accelerometer sensor and the GPS embedded in a smartphone to interact directly with the software. MYS is based on an existing mobile application called Walking the Edit (http://walking-the-edit.net) where you can "*walk your movie*".

The Walking the Edit application generates a movie automatically according to the position and the profile of a user. It searches for the best existing videos (called assets) and combines them in order to produce a new coherent content. These assets are short videos referenced by their localisation and meta-data. There are three applications domains depending on the expectations and needs of the end user:

- Touristic applications
- Artistic / cultural projects
- Historic / patrimonial use cases

The system searches for all assets related to the context and preferences of the user and then mixes them together automatically. The user can then publish and share the personalized movie(s) on social platforms (Facebook, Twitter, etc.), and can also contribute with his own content "on the way" or afterwards, through a simple editing interface.

The MYS project is aimed at improving this application by taking into account the current behaviour of the user in the assets

This work was supported by the Swiss Commission for Technology and Innovation (CTI grant 13943.1)

selection process. This is done by detecting activities such as walking, running, bicycling or travelling in a vehicle.

The proposed activity recognition engine combines the GPS and the accelerometer embedded in the mobile device. This way, it can recognise the current activity and its level when the user carries his device in the hand or in a pocket in any orientation. Furthermore, it works both indoor and outdoor. Section II describes our approach for detecting the activity. The user study validation and its results are presented in Section III and the interaction with the editing engine in Section IV. Finally, Section VI presents a conclusion and future work.

II. METHODOLOGY

The MYS project is composed of two main parts: the activity recognition module called BehaviourMod (presented in this section) and the editing engine which selects the right part of movie related to the user's activity (cf. Section IV).

BehaviourMod processes data from the accelerometer sensor and the speed computed from the GPS signal. The Editing Engine requires the current user's activity before selecting the next asset. Therefore, BehaviourMod needs to run continuously in order to provide accurate information about the current activity.

The MYS application can be used without constraints, in a pocket or in hand. In this case, the activity is detected with the arm moving along the body or with the hand holding the device when the user is looking at the screen. Finally, the application continue to perform activity recognition when the GPS signal is lost, e.g. when the user is inside or closed to a building.

A. Acceleration processing

Acceleration data are sampled at 40 Hz from the 3-axis accelerometer sensor embedded in an Android mobile phone. Specifically, the sensor which provides acceleration information without the gravity component (linear accelerometer) is used.

The norm of each sample vector is calculated, and then low-pass filtered in order to remove some of the noise. The signal is also scaled by 40%. Finally, an acceleration density (AD) factor is computed as the median of the norms collected over 2 seconds. The *AD* represents the activity intensity in m/s^2. This approach is similar to the activity counts described in [8] and detailed in [9].

B. GPS speed processing

We are using the GPS embedded in the smartphone to get the current speed of the user. The sampling interval is set to 1 second. The speed is directly provided by the GPS. In order to remove noise from the signal, we average the last 3 values. The GPS signal is used only when the error is equal or inferior to 20 meters.

The *AD* value is principally used when the device follows the user movement, e.g. positioning in a pocket. If the user is walking and holding the device in his hand (e.g., looking at something on the screen), the *AD* does not reflect the real movement intensity. In this case, the GPS can compensate with the speed value.

C. Experimentations

Several experimentations in different conditions were done in laboratory settings and in real conditions. The following graphs show typical examples of signals recorded during selected activities. For visualization purposes, we decided to plot the *AD* factor (m/s^2) and the GPS speed (km/h) on the same graphs.

Fig. 1 represents the *AD* factor and the GPS speed sampled every 2 seconds of a user walking normally on a street. The phone was in the user's front trouser pocket. The *AD* curve showed a regular walk with values in a range from 13 to 20 m/s^2. The GPS speed confirmed the walking activity with a range from 1 to 6 km/h and an average at 4.67 km/h. We observed some low values in the GPS speed which were caused by the low accuracy of GPS at low speeds.

The *AD* curve in Fig. 2 shows a clear difference between a bicycling activity and a walking activity. In the two previous examples, the *AD* for bicycling is on average at 3.57 m/s^2 and at 16.22 m/s^2 for walking activity. This difference can be explained by the movement itself. When bicycling with the device in a pocket, the movement is very smooth. However, while walking, the movement is more chaotic, mostly because of the impact of the feet on the ground. Like for the walking example, the GPS speed value reflected again the right activity. Except for the running activity, a walk cannot reach the speed (average at 14.13 km/h) depicted in Fig. 2.

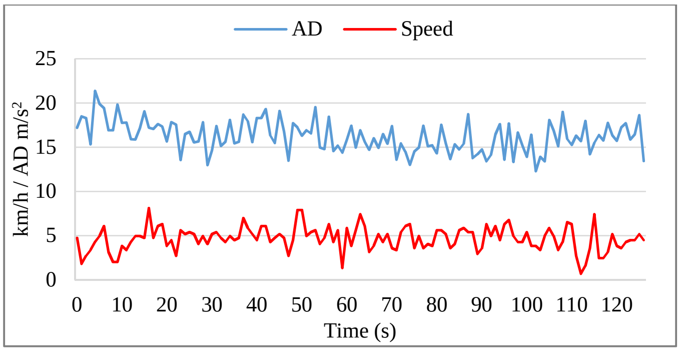

Figure 1. Walking activity.

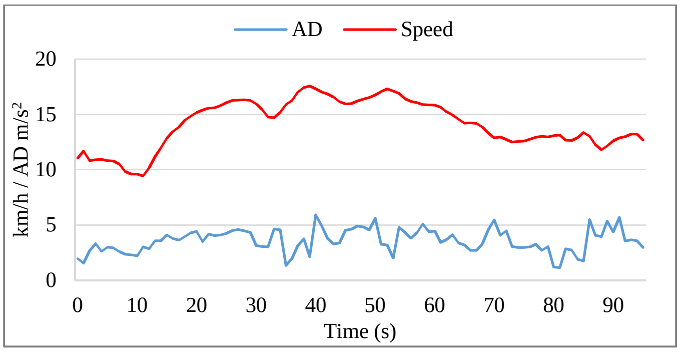

Figure 2. Bicycling activity.

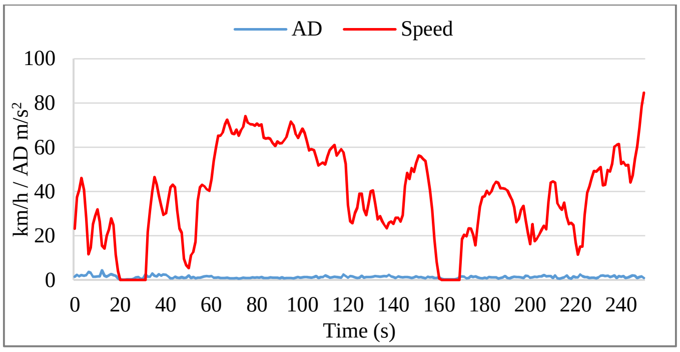

Figure 3. Driving activity.

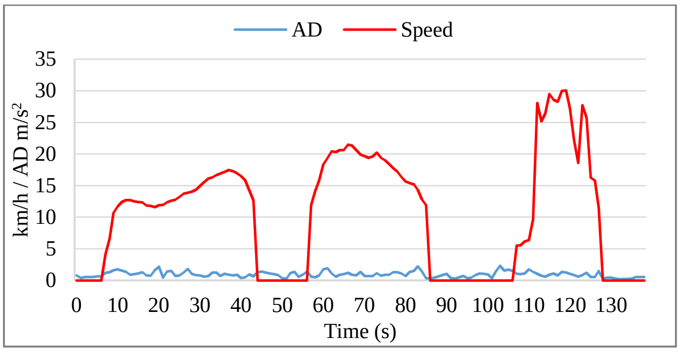

Figure 4. Travelling in a tram.

While driving (Fig. 3), the AD remained low with an average at 1.33 m/s^2. We found some signal peaks while the user was accelerating. The phone was in the user's front trouser pocket.

Movements while travelling in a tram (Fig. 4) are similar to the ones shown in Fig. 3. The *AD* is slightly lower and more regular in this case with an average value of 0.94 m/s^2.

By comparison, the *AD* factors for driving, travelling in a tram and bicycling are very similar. The GPS speed is the principal value for making the distinction between bicycling and the other activities. However, in the case of travelling with low speed (e.g. stuck in a traffic jam) a misclassification can occur. In such cases, the AD factor can help to predict the correct activity.

D. *Activity Classification*

We propose a decision tree classifier to detect the current activity. Using the computed *AD*, we classified the activity into 3 domains: static, walking and "other activities". The walking domain included 3 walking intensity levels and the running activity. The "other activities" domain included bicycling and travelling in a vehicle. Experimentations showed that the *AD* was always much higher for walking than for other activities.

In the walking domain, we set 4 thresholds for defining the activity into 3 walking activity levels: slow, normal, fast and the running level. These levels were related to the walking speeds levels (Table I) defined in [10] and calibrated with a user study (cf. Section III).

TABLE I. SPEED RANGE

Level	Speed (Km/h)
Slow	1 – 3
Normal	3 – 5
Fast	5 – 7
Running	> 7

TABLE II. MINUTES COLLECTED PER ACTIVITY

Walking	On Vehicle	Bicycling	Static
115	186	84	29

TABLE III. DATA AVERAGE PER ACTIVITY

	Walking		On Vehicle		Bicycling	
	AD	*Speed*	*AD*	*Speed*	*AD*	*Speed*
Average	16.42	4.82	1.38	43.96	3.81	14.79
Std.	3.56	1.15	0.52	35.53	1.16	3.58
Max	22.91	7.98	2.99	128.12	6.14	21.72
Min	5.63	1.13	0.24	5.18	1.02	2.94

The second domain gathered together other activities like bicycling and travelling in a vehicle. The activities were classified by combining the *AD* with GPS speed. Finally, a static activity was defined for speeds under 1 km/h and *ADs* under 1 m/s^2. We used a tolerance to avoid recognizing activities when the user is interacting with the device without moving.

The decision tree classifier was calibrated with data collected from 5 different users for a total duration of approximately 7 hours (12'443 samples). Participants were asked to collect data while commuting from home to their office. A custom mobile application was developed for this purpose.

All participants used different transportation modes during this trial. All displacement were labelled by the user directly in the graphical interface of the application.

Table II summarises the durations in minutes collected per activity. The resting time was labelled as a static activity. We also combined the driving activity with the travelling in a vehicle activity as their data were very similar.

Table III presents the average, standard deviation, minimum and maximum values of all collected *AD* (m/s^2) and GPS speed (km/h) for each activity.

Based on the result presented in Table III, we refined the thresholds used in our decision tree classifier (Fig. 5).

E. *Remarks*

While testing our algorithm, we found some possible errors on the bicycling activity. The main cause of such errors was poor GPS accuracy. For example, when a user is walking slowly (low *AD*) and if the GPS speed is too high (due to bad accuracy), the algorithm will detect a bicycling activity. The inverse misclassification was also found when bicycling at a low speed. The *AD* helps detecting the right activity in this case. We also added a GPS speed limit to avoid the possible misclassification with the in a vehicle activity.

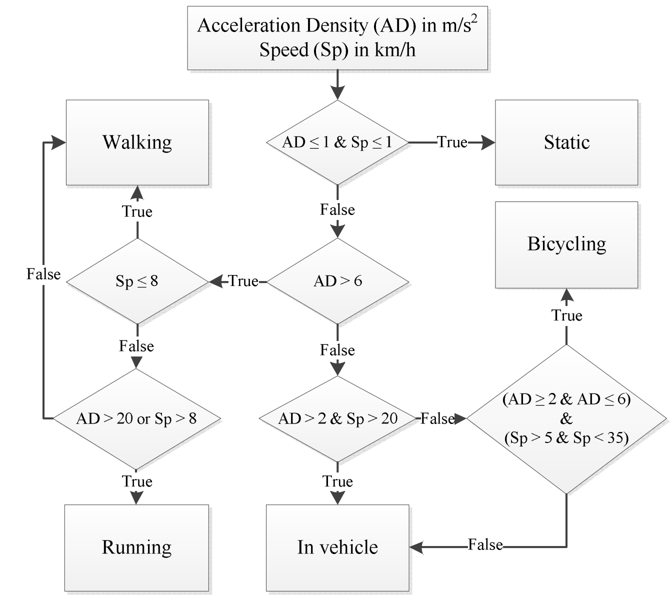

Figure 5. *Decision tree classifier*

When travelling in a vehicle, we observed that most of the time, the *AD* was really low. Some exceptions can occur during quick accelerations, but in this case, the GPS speed is used to select the correct activity. Finally, another issue can occur when the user is in a vehicle moving slowly and with some bad road conditions. In this case, in order to avoid detecting a bicycling activity, a rule that prevents switching from a in vehicle activity to a bicycling one and vice-versa repeatedly is used.

III. VALIDATION

A user study was carried out for validating our algorithm. 5 people participated in the study, 3 males and 2 females.

Data recorded were also used to calibrate the 3 walking levels and the running activity that were pre-defined.

A. *Protocol*

All participants performed the following steps (2 minutes per activity):

- Walking (3 – 5 km/h)
- Fast Walking (> 5 km/h)
- Slow Walking (< 3 km/h)
- Bicycling
- Running

TABLE IV. VALIDATION STUDY RESULTS

	Activity detected (%)		
Steps	*Walking*	*Running*	*Bicycling*
Walking	100	0	0
Walking fast	95.4	4.6	0
Walking slowly	94.8	0	5.2
Bicycling	32	0	68
Running	5.4	94.6	0

TABLE V. BICYCLING RESULTS WITH 4 USERS

	Activity detected (%)		
Steps	*Walking*	*Running*	*Bicycling*
Bicycling	12.5	0	87.5

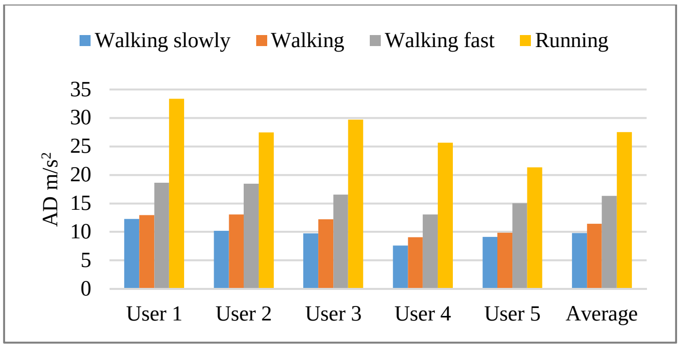

Figure 6. *AD variations intensity per user.*

All participants carried the device (Galaxy Nexus) running a custom application for data logging in their trouser front pocket. The acquisition sessions for all subjects were recorded on video in order to have a reference to annotate the data.

B. *Data synchronization*

For each participant, we took the start and end times of each step of the protocol in the video. Then, we processed data (about 60 data samples per step) matching with each period. Finally, we assessed the performance of the system by computing the ratio of the number of correctly classified samples to the total number samples for each activity.

C. *Results*

Table IV summarises all results of this validation study. For each step of the protocol, we present the average accuracy of all detected activities.

Overall, the walking activity was correctly detected. When walking fast, the algorithm detected a running activity for short durations because the *AD* was at the limit of the threshold discriminating walking and running activities. The same comment is relevant for the running activity.

On the slow walking step, the *AD* was very low and the GPS speed was inaccurate. One user was too close to a building during the trial. Data recorded during this step showed a speed around 10 km/h in the first 30 seconds and therefore, the bicycling activities detected were false positives.

We had difficulties during the bicycling step. During the bicycling activity of the first user, the GPS signal was lost and thus no speed was detected. Consequently, the whole step was detected as a walking activity. Table V presents the average bicycling results with only 4 participants:

D. *Walking level calibration*

We used data recorded in this study to refine our *AD* – GPS speed thresholds for the walking levels. Fig. 6 presents the *AD* variations intensity per walking level and per user.

TABLE VI. WALKING LEVEL DEFINITIONS

	Walking slowly	Walking	Walking fast	Running
AD	< 11	≥ 11 & < 15	≥ 15 & < 20	≥ 20

TABLE VII. WALKING LEVEL RESULTS

	Average walking level detected (%)		
Steps	**Slow**	**Normal**	**Fast**
Walking slow	78.6	21.4	0
Walking	17.8	72.6	9.6
Walking fast	0	38	62

The differences in AD were large enough to be detected by our algorithm. We implemented a second decision tree classifier under the walking domain. This classifier is working in the same way than the one described in Section II. Each walking level is defined by thresholds composed of AD and GPS speed. We discovered that the AD was not sufficient to categorise the walking level as the device might be held in the user's hand. In this case, the AD alone cannot be used to determine the correct walking level. Towards this end, we compared the GPS speed in Table I and the AD intensity in Table VI. We used the same data set for calibrating the decision tree classifier. Table VII summarises the classification ratio of the walking level correctly detected.

IV. MYS EDITING ENGINE

The "editing engine" is an algorithm that dynamically creates a playlist of media (assets) based on the analysis of the walk of a user. This process is performed at the end of each currently playing asset. This special search engine is not based on words or values, but instead on the user's behaviour over a period of time. BehaviourMod influences the editing engine in multiple ways.

The first influence is on the assets pre-selection. As the MYS application contains a multi-layer mixing system, which combines narrative, effects and background assets at the same time. The engine selects completely different assets depending on the user's behaviour and creates an entire environment based on the user's position and behaviour. This is done thanks to the meta-data contained in the assets and it allows the editing engine to filter those which will be displayed while walking, running, bicycling etc.

The second influence regards the assets scoring system. At time of writing, the MYS database contains more than 1800 assets geolocalised in the city of Geneva. A group of the 100 assets closest to the user position is defined, the engine then computes a score for each asset and determines which one to play next. This score is computed in the editing engine and has several parameters such as the distance between the user and the asset, the movie theme, the asset theme, the user's speed, and the user's behaviour.

A. Scoring algorithm

All scoring values were determined in a heuristic way. These values can all be changed depending on which parameters should have more influence on the asset selection process.

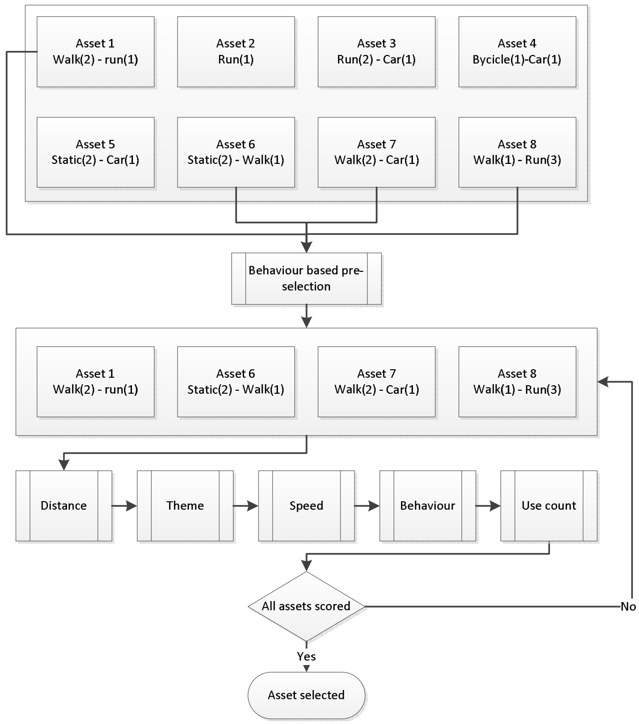

Figure 7. Asset selection process.

The following example (Fig. 7) presents, in a simplified way, how the editing engine selects the different assets. We suppose that the detected activity is "walking". We also assume that the movie theme is "culture" and that the user is moving at 5 km/h.

Only 4 assets are selected based on the walking behaviour (asset N°: 1 - 6 - 7 - 8). These assets then go through the scoring algorithm which increases their score depending on whether they fit the user's context/behaviour. Each of these parameters has a bonus weight.

The score of each asset is initialized based on its distance to the user. The algorithm selects the 100 closest assets. The closest one is assigned a 200 points score. The score assigned to each subsequent asset is decreased by 2 points.

Asset N° 6: score = 200

As the theme of the asset fits the theme of the ongoing movie, the asset obtains a bonus of 50 points.

Asset N° 6: score = 200 + 50

The influence of speed is derived from a mapping table between speed and asset duration, which allows the engine to attribute more points to the fitting assets (short assets correspond to high speed). Depending on the asset duration and the current speed, a full bonus is selected (50 points) or it is divided by a factor given in Table VIII.

Assuming that the Asset N° 6 has a length of 37 seconds and the user is moving at 5 km/h, according to Table VIII, the speed bonus is divided by 2.

Asset N° 6: score = 250 + 25

TABLE VIII. SPEED MAPPING – DIVIDER FACTOR TABLE

Speed (km/h)	Duration (s) 45 to ∞	30 to 45	15 to 30	10 to 15	0 to 10
< 3	-	2	3	4	5
≥ 3 & < 5	2	-	2	3	4
≥ 5 & < 7	3	2	-	2	3
≥ 7 & < 15	4	3	2	-	2
≥ 15	5	4	3	2	-

TABLE IX. ASSETS CHARACTERISTICS

	Assets details			
	Asset N° 1	*Asset N° 6*	*Asset N° 7*	*Asset N° 8*
Distance to the current position	37 m	12 m	55 m	34 m
Theme	Culture	Culture	History	Culture
Duration	58 s	37 s	89 s	8 s
Behaviour tag	Walk(2) Run(1)	Static(2) Walk(1)	Walk(2) Car(1)	Walk(1) Run(3)
Use count	1	0	0	0

TABLE X. ASSETS SCORES TABLE

	Assets scores[a]			
	Asset N° 1	*Asset N° 6*	*Asset N° 7*	*Asset N° 8*
Distance rank	196	200	194	198
Movie Theme	50	50	0	50
User Speed	50/3	50/2	50/3	50/3
User Behaviour	25*2	25*1	25*2	25*1
Use Count	-50	0	0	0
Total	263	300	261	290

a. We assume that the theme selected is Culture, the current speed is at 5 km/h and the current behaviour is walking.

The behaviour parameter is slightly more complex as each asset contains several behaviour tags. At this stage of the process, each pre-selected asset contains the "walking" behaviour tag. (Pre-selection was made previously). Each of these tags has a weight. For example, "Asset N° 6" is more fitted for static but can still be displayed while walking. Therefore, it gets a weight of 2 for static and a weight of 1 for walking. This weight is a multiplicative factor for the behaviour bonus set at 25 points and added to the score.

Asset N° 6: score = 275 + 25*1

The last step gives a negative bonus if the asset was already played: 50 points multiplied by the playing counter the asset was played are removed from the score. This functionality was set in order to have a continuous asset turnover. After playing an asset, a tag called use-count representing the number time it was played is incremented. In this example we assume that no asset has been played before.

Asset N° 6: score = 300 – (50 * 0)

Finally, the Asset N° 6 is given the score of 300 points. This procedure is done for the 100 closest assets, and the one with the highest score is selected to be played next as it is the most fitting asset, taking the context and the user's behaviour into account.

Table IX presents characteristics available for asset samples. Individual scores for these assets are presented in Table X.

V. CONCLUSION

The smartphone application called MoveYourStory generates a coherent movie based on a selection of many small videos clips called assets. The assets are referenced by meta-data and geolocalised into a database. The selection of these videos is done by an editing engine which translates the user's behaviour into a story. By this way, the user is able to influence with his body the movie narration by its own behaviour. By walking at different speed, or suddenly running, the editing engine will select the right piece of video and adding it to the playlist. Thanks to this interaction, each generated movie is unique.

The behaviour is estimated with an activity recognition algorithm and applied to interact with MYS. We implemented a decision tree classifier to detect activities based on the accelerometer data and the GPS speed together. The decision tree classifier was tested with a user study and the recognition accuracy for walking activities was 96.7% and 87.5% for the bicycling activity. A deeper classification was also calibrated for detecting walking levels.

New features are already planned for the MYS project. In order to improve the user's experience with this application, we will experiment how to detect another set of behaviour patterns. By adding the orientation to the running process, we expect to detect – on a time scale of 30s – if the user seems lost or stressed by moving in a chaotic way. We also plan to detect gestures when using the smartphone. All these new features will be used in the new asset selection process, but also to add sound effects to the generated movies.

REFERENCES

[1] V. Stewart, S. Ferguson, P. Jian-Xun, and K. Rafferty, "Practical automated activity recognition using standard smartphones," in Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on, 2012, pp. 229-234.

[2] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," *Ubiquitous Intelligence and Computing*, pp. 548-562, 2010.

[3] M. Fahim, I. Fatima, S. Lee, and Y.-T. Park, "EFM: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer," *Applied Intelligence*, pp. 1-14, 2013.

[4] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, "Sensing meets mobile social networks: the design, implementation and evaluation of the CenceMe application," in Proceedings of the 6th ACM conference on Embedded network sensor systems, 2008, pp. 337-350.

[5] P. P. Jayaraman, C. Perera, D. Georgakopoulos, and A. Zaslavsky, "Efficient Opportunistic Sensing using Mobile Collaborative Platform MOSDEN," *arXiv preprint arXiv:1310.4052*, 2013.

[6] S. Hemminki, P. Nurmi, and S. Tarkoma, "Accelerometer-based transportation mode detection on smartphones," in Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, 2013, pp. 13.

[7] T. Feng, and H. J. Timmermans, "Transportation mode recognition using GPS and accelerometer data," *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 118-130, 2013.

[8] K. Y. Chen, and D. R. Bassett Jr, "The technology of accelerometry-based activity monitors: current and future," *Medicine and science in sports and exercise*, vol. 37, no. 11 Suppl, pp. S490, 2005.

[9] J. Hausmann, and K. Wac, "Activity Level Estimator on a Commercial Mobile Phone: A Feasibility Study," in International Workshop on Frontiers in Activity Recognition using Pervasive Sensing (IWFAR), co-located with the PERVASIVE Conference, San Francisco, CA, USA, 2011.

[10] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Bassett Jr, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon, "2011 Compendium of Physical Activities: A Second Update of Codes and MET Values," *Medicine & Science in Sports & Exercise*, vol. 43, no. 8, pp. 1575, 2011.