20th International Conference on Information Fusion
Xi'an, China - July 10-13, 2017

Motionword: An Activity Recognition Algorithm based on intelligent terminal and cloud

Zhenjie Yao*†, Zhipeng Zhang†, Junyan Wang†, Li-Qun Xu†
*Beijing advanced innovation center for future internet technology, Beijing University of Technology, Beijing, China
Email: yaozhenjie@gmail.com
†Center of Excellence for mHealth & Smart Healthcare, China Mobile Research Institute, Beijing, China

***Abstract*—The ability to recognize physical activity, such as sedentary, driving, riding, daily activities and effective training, is useful for health conscious users to catalogue their daily activities and to develop good exercise routines. Conventional activity recognition algorithms require complex calculations, which are not suitable for wearable devices developed on low-cost, low-power hardware platforms. In this paper, inspired by the text mining related work, we design a novel activity recognition algorithm, which is named "Motionword". In the wearable device proper, a lightweight recognition algorithm is adopted to compute in real-time predefined atomic events, and count the frequency that these events occur, resulting in a data summary, and then the data summary is transmitted to the platform. On the platform, intelligent method is used to identify and categorize the user's main activity into 5 classes. The test results on a dataset composed of 110 user*day real world data, contributed by 10 users, show that the recognition accuracy is 95.52%. The Motionword algorithm is capable of achieving accurate activity recognition results without additional hardware cost or power consumption.**

I. Introduction

Regular physical activity improves heart and lung function; reduces the incidence of Non-communicable Diseases (NCDs) and health risk factors, such as coronary heart disease, high blood pressure, high blood sugar, fat metabolism disorders etc.; enhances bone strength, slows the progress of osteoporosis; reduces stress, maintains mental health and prevents depression. In addition, it can also reduce obesity and help one to keep fit [1][2]. On the contrary, Physical inactivity is the main cause of breast or colon cancers, diabetes and heart disease, and it is the fourth leading risk factor for global mortality (6% of global deaths) [3].

There are a lot of wearable devices (Fitbit, Nike+, Jawbone, Omron pedometers, etc.) with step counting function. However, these devices can only count the number of steps, and make a rough estimation of distance and energy consumption based on the number of steps. These devices cannot identify the user's physical activity. User activity recognition enables fully health monitoring [4]. Scientific directions and guidance can be delivered to the user based on the activity log. Smart phone is claimed to be an ideal platform for data mining applications for its small size, computation power, communication capabilities and ubiquitous use in our society [5][6]; however, we do not think so. Generally, when you are in office or at home, the smart phone lies on the desk, or charging. Moreover, carrying smart phone during exercising is annoying. We still think that wearable devices are more suitable for successive activity recognizing. It is necessary to achieve an activity recognition algorithm that is suitable for wearable devices.

However, the existing activity recognition algorithm is carried out based on the raw data from accelerometer and/or gyroscope, with feature extraction and classification [7][8][9]. These algorithms are not affordable for the current wearable devices that developed with low cost, low power hardware platforms.

There are many data summary extracting technologies in text mining domain [10][11], such as the vector space model based extraction techniques, like the famous TF.IDF (Term Frequency. Inverse Document Frequency). With this technology, we can represent an article with thousands of words by key word frequency. This technique has been extended to the field of visual recognition [12][13], named Visualword algorithm. Inspired by the summarizing technology, we designed a new algorithm to recognize user's activities, which is named Motionword. The algorithm takes advantage of both cloud platform and intelligent terminal. In the terminal, a lightweight algorithm processes the data and summarize it into a "motion frequency" vector. The "motion frequency" vector is transmitted to the platform, where the intelligent KNN (K Nearest Neighbor) classifier identifies activities into five categories: sedentary, driving, riding, daily activities and effective exercise.

The remainder of this paper is organized as follows: section II gives the architecture of the whole activity recognition system. The principle of Motionword algorithm is introduced in detail in Section III. Test results on a practice dataset are shown and discussed in Section IV. Finally, we conclude this paper and discuss future work in Section V.

II. Activity recognition system architecture

The architecture of the whole activity recognition system is shown in Fig. 1, the user wears a pedometer, with a low-power MSP430 processor, which has a memory of only 6k and a processing capacity of 25MIPS (Million Instructions Per Second). The battery capacity of the system is only 200 mAh. Due to the limitation of memory, processing capacity and battery capacity, it is impossible to process long data with complex algorithm, only lightweight processing of short time data is feasible.

As shown in Fig. 1, a lightweight recognition algorithm is adopted to compute in real-time predefined atomic events, and count the frequency that these events occur, resulting in a data summary, and data summary is uploaded to the platform.

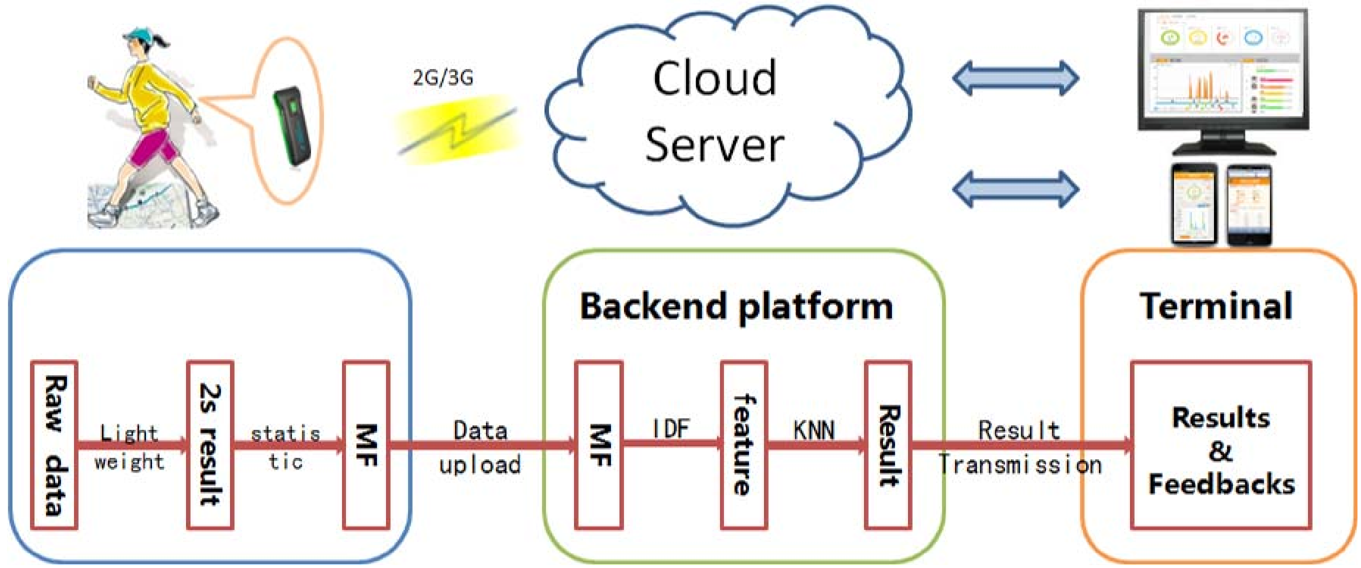


Fig. 1. The architecture of the whole activity recognition system.

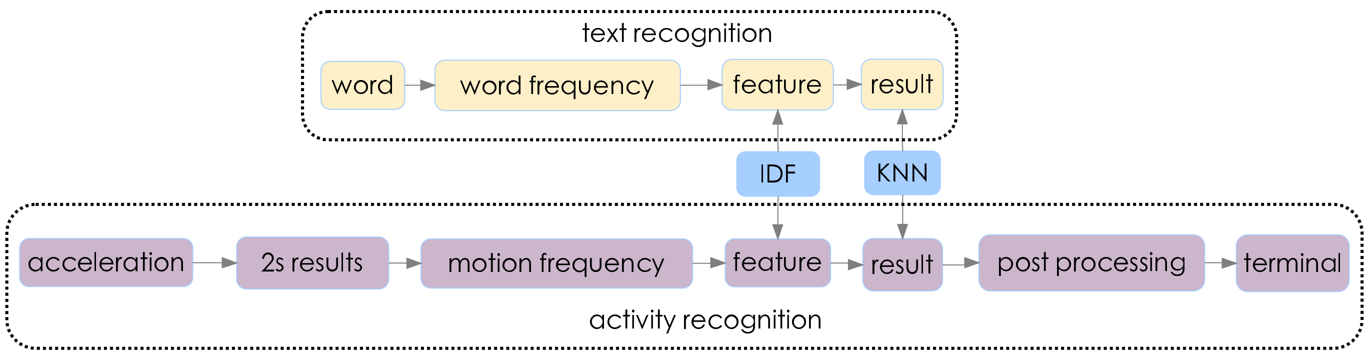


Fig. 2. Comparison between Motionword and text mining.

The platform is responsible for receiving, parsing and storing the data. When the user the activity recognition results via our website or via our application on the phone, a request will be send to the platform, the platform gets the data, and calls an intelligent algorithm to identify the 5 kinds of physical activities. Then the platform pushes the final results to the terminal. Personalized guidance that deduced from individual activity recognition results are also included.

III. MOTIONWORD ACTIVITY RECOGNITION

This section describes the implementation details of Motionword algorithm. For ease of understanding, we present each step and compare it with a corresponding step of text recognition algorithm, as shown in Fig. 2.

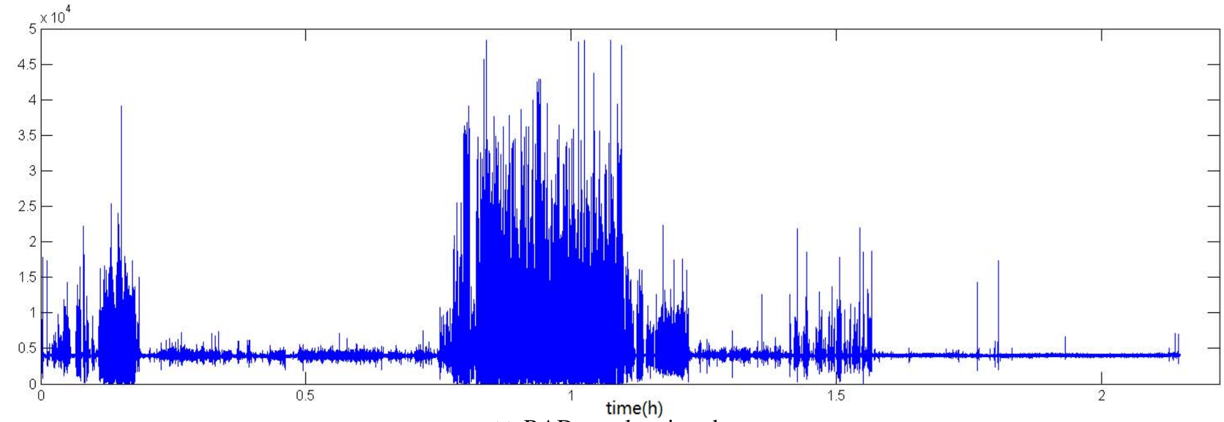
A. *Intelligent terminal algorithm*

First, at the terminal, here it is a pedometer, in order to reduce the memory usage, we cannot store raw data for a long time, decision should be given based on a short piece of data. Here, we recognize the motion every 2s based on the past 5s. For each piece of data, we calculate the sum of the square of triaxial acceleration. Denote triaxial raw data as a_x, a_y, a_z, and the sum of the square is:

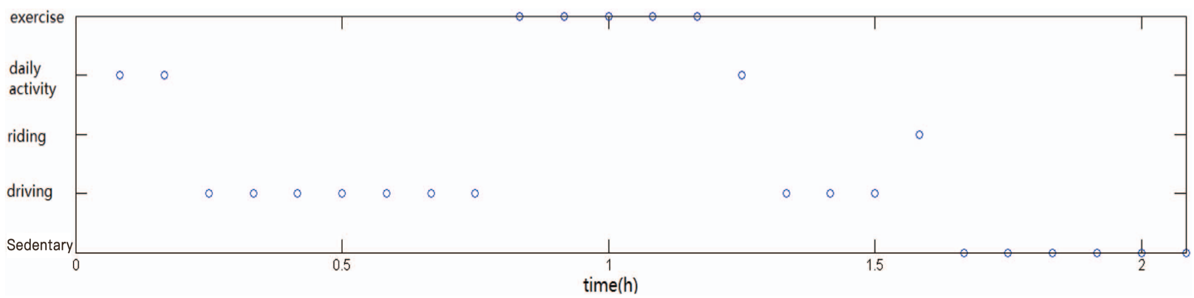
$$f = a_x^2 + a_y^2 + a_z^2. \quad (1)$$

By observing actual acquired data, we found that the variance $var(f)$ of different activities usually meet the following rules (2).

According to this rule, we can identify the current state of motion based on the variance. On the other hand, the variance calculation is affordable on the hardware platform. We use variance for preliminary identification of the atomic activities. First of all, we determine the classification threshold corresponding to different activities through statistical $var(f)$, then we could identify each piece of data into one class (sedentary, driving, riding, activities, exercise). The atomic result of each piece corresponds to a "word" in the text recognition. A "word" sequence (i.e., the result of long sequences) can be regarded as a "text" (here data of five minutes as a "text", its length is 150). Next, we summarize the text and statistic the frequency of occurrence for each "word" as feature. For every 5 minutes, we obtain a "motion frequency" vector of 5-dimensions. For example, "motion frequency" [20,10,5,45,70] represents that, in the 5 minutes, 20*2s periods are identified as sedentary, 10*2s periods are recognized as driving etc. This is a typical "motion frequency" vector during strenuous



(a) RAD acceleration data



(b) RAD recognition result

Fig. 3. Example of RAD dataset.

$$var(f)_{sedentary} < var(f)_{driving} < var(f)_{riding} < var(f)_{activities} < var(f)_{exercise}. \tag{2}$$

exercise. "Motion frequency" vectors are transmitted to the platform directly. "Motion frequency" is corresponding to "word frequency" in text recognition.

B. Intelligent Recognition on cloud

Back-end cloud server accesses the "motion frequency" message, combined with the IDF. IDF is defined as (3)

$$IDF_i = \frac{|D|}{|\{j : t_i \in d_j\}|}. \tag{3}$$

in which $|D|$ is the total number of documents, here each 5 minutes is a document. d_j is the jth document. t_i is the ith word, it is a kind of activity. In fact, IDF is actually a weighting strategy according to the importance [14], weights corresponding to different activities vary. Almost every type of activity may include "sedentary". For example, playing basketball is a strenuous exercise, but during 5 minutes, there may be transient pauses, some of total 150*2s periods are identified as sedentary. Therefore, containing 60s sedentary does not characterize the overall activity in this 5 minutes. That means the contribution of the sedentary to activity recognition is small, and the corresponding IDF weight should be small. On the contrary, exercise rarely appears during other activities (sedentary, driving, riding or activities). If there are 60s of exercise within 5 minutes, then it is certain that the user is taking exercise during the 5 minutes. Therefore, the contribution of exercise to activity recognition is great, and the corresponding IDF weight should also be large.

Weighted "motion frequency" is the final feature vector that fed in KNN (K Nearest Neighbor) classifier, which takes cosine distance (3) to measure similarity. The recognition result is given every five minutes. Cosine distance is defined in (3). Where θ is the angle between the two vectors X and Y.

The recognition results may contain occasional errors. We can correct a part of these errors according to the experience of life. For example, if a 5-minute strenuous exercise occurs abruptly in a number of daily activities, it is a error with high probability. It can be corrected. If a 5-minute driving bursts during a long time of being sedentary, it is probably a misclassification, and can be corrected, too.

IV. Experimental Results

In order to verify the effectiveness of the algorithm, it was tested on the two datasets. One is Raw Acceleration Dataset (RAD), the other is Intermediate Results Dataset (IRD). Both data sets are labeled by the user manually.

Both datasets are captured by 10 users, 4 females and 6 males. The users aged between 25 to 45. It worth to mention that we did not restrict the wearing position, users put the device on the waist or in the pockets (including jacket pockets and trousers pocket) at their convenience.

RAD is composed of raw acceleration data collected by an customized sensor nodes; the data was transmitted to the cell phone via Bluetooth for recording. Constrained by node

$$Sim(X,Y) = \cos(\theta) = \frac{x_1y_1 + x_2y_2 + \cdots + x_ny_n}{\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \cdot \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2}}. \quad (3)$$

power, each acquisition lasted about 2.5 hours until the battery (200mAh) runs out. 10 users participated in the collection; a total of about 40 hours of raw acceleration data was collected.

IRD is composed of intermediate results send to the platform in the actual application scenarios. Data throughout the day is captured. The battery capacity is 200mAh, the wearable device can work about 14 days. The dataset includes data from 10 users, each with 11 days, totally 110 user*day of data. The recognition unit is 5 min; there are 288 recognition units per day. Therefore, the test dataset is composed of 288*110 recognition unit.

Fig. 3 is an analysis result of data in RAD database, wherein Fig. 3(a) is the acceleration data, which can assist the user as a prompt for data annotation. This acquisition lasted 2 hours and 5 minutes. The truth (annotated by the user) is that: the user spent 10 minutes tidying the clutters and walking to the company parking lot, 35 minutes driving to the basketball field, 25 minutes playing, and then 5 minutes walking around. Afterwards, 15 minutes driving to the parking lot, 5 minutes walking to the house, and then keeping still. Fig. 3(b) is the analysis results of Motionword. The results are almost entirely consistent with annotations. One period at 1.5 hours, which is the last activity, is the only error. The activities within this 5 minutes are very complicated, including parking, walking, riding the elevator and sitting at home. Due to the activities in this 5 minutes containing several state transitions, the transition point state (i.e., switches from one state to another state period) will cause preliminary recognition error on nodes, and identified as riding on the back-end server. Test results on RAD show the recognition accuracy rate reached 94.21%.

Fig. 4 is an analysis result of data in IRD. Fig. 4(a) is the step number of the user. The step counting error of the pedometer is very small, which can be used as prompts to assist the user for data annotation. Fig. 4(b) is the analysis results of Motionword. The results are almost entirely consistent with annotations.

Table 1 presents the confusion matrix on IRD, detailed statistical analysis is shown. The overall recognition rate is 95.52%. The main error occurred in driving, riding and effective exercise. The recognition results of driving and riding are related to speed. By observing actual data, we find that, if the speed is faster, or the road bumps, driving is easier to be mistaken for riding or daily activities. By the same token, if the user rides slowly, riding is easily recognized as driving. On the contrary, if the user rides fast, riding is easily recognized as daily activities. Part of the daily activities is confused with the effective exercise, which is caused by transitional state. Besides, some official activities and housework (cooking, laundry, playing with children, etc.) tends to be mistaken for driving or riding. The reason is that when doing the housework, the body will have a small amplitude swing, and there is little margin between housework body shaking and shaking produced by driving a car or bicycle.

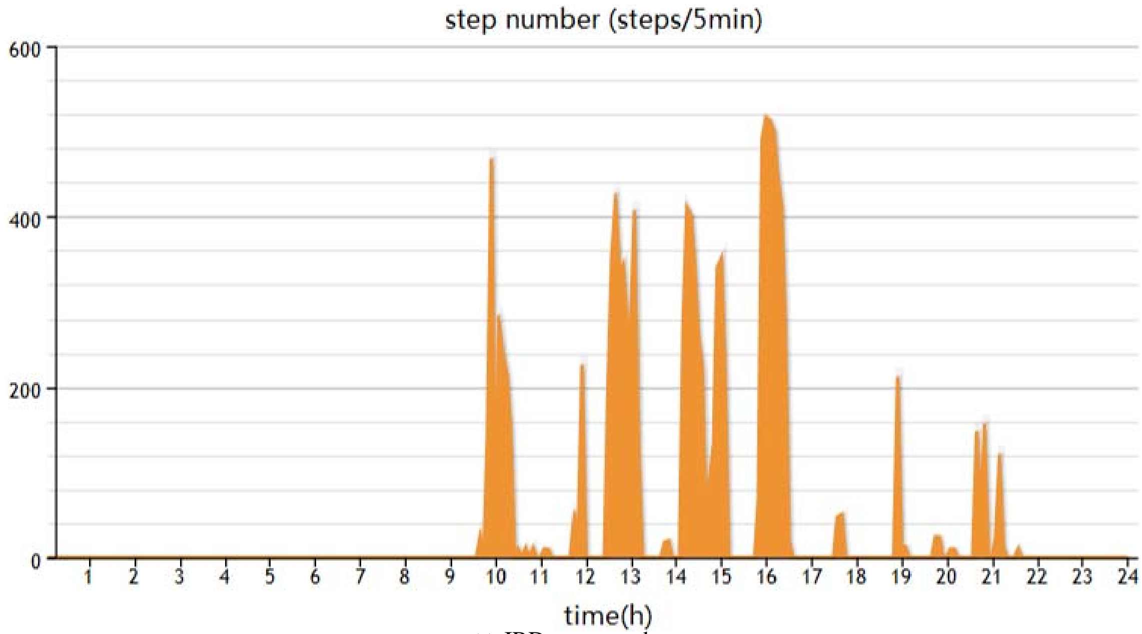
V. CONCLUSIONS

In this paper, by imitating data summary extracting technologies in text mining domain, we designed a novel activity recognition algorithm, which is named Motionword. In the intelligent terminal, a lightweight recognition algorithm is adopted to get preliminary results, counting the frequency of the initial results, induces a data summary. Then the data summary is uploaded to the back-end cloud platform, where intelligent method (KNN) is used to identify the user's main activity into 5 classes: sedentary, driving, riding, daily activities and effective exercise. 110 user*day of actual test data demonstrated the feasibility and effectiveness of this method. On a low-cost, low-power embedded hardware platform, the final overall recognition accuracy is as high as 95.52%. To achieve a more accurate analysis of user activity, current problems and further efforts include:

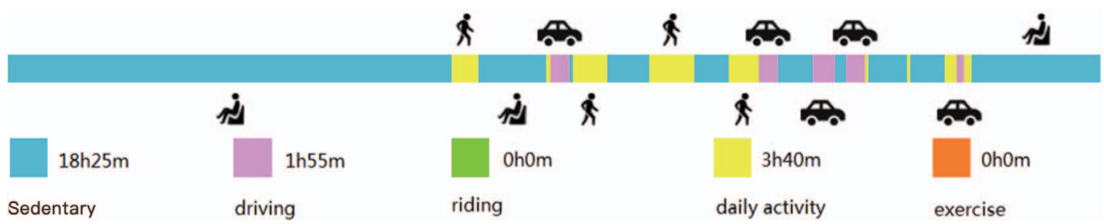
i. Motion frequency need to be improved. Now the sequential information is lost. adding a few dimension to the motion frequency vector may improve the recognition accuracy;
ii. Office activities and home activities (cooking, laundry, playing with children, etc.) are easily mistaken for driving or riding. This is because there will be slight body shaking when doing housework, which is similar with driving or riding;
iii. Riding tends to be misclassified; riding speed and wearing position have significant influence on the activity recognition results. The recommended wearing position is on the waist belt;
iv. When the user is driving, the MCU is possible to go to sleep mode. The waking up threshold should be determined elaborately, which is a tradeoff between recording time and energy saving;
v. Location aided activity recognition. The test results will be mainly concentrated in the error driving/riding identified as sedentary or moving. Generally, driving/riding occur simultaneously with rapid location change, the location information may be used to improve the activity recognition.
vi. Guidelines based on activity recognition results. According to WHO guidelines [1], adults should participate in exercise for 30 minutes per day, or at least 150 minutes per week. We can obtain the activity category and duration after recognition, from which we can deduce the amounts of each kind of physical activity, the times and duration of sedentary. If the user did not do enough activities as recommended, we can recommend the user to join a specific length of activities to achieve fitness goals. Users would also be warned if he/she keep sedentary for long time, e.g. two times of sedentary per day, he/she should avoid any sedentary.

REFERENCES

[1] W. H. Organization *et al.*, "Global recommendations on physical activity for health," *Geneva: World Health Organization*, pp. 8–10, 2010.



(a) IRD step number



(b) IRD recognition result

Fig. 4. Example of IRD dataset.

TABLE I. CONFUSION MATRIX ON IRD DATASET

Annotation result	Sedentary	Driving	Riding	Daily Activity	Exercise	Accuracy
Sedentary	26872	690	82	197	13	96.47
Driving	23	944	63	38	5	87.98
Riding	1	0	32	9	0	76.19
Daily activity	45	22	29	2054	198	87.48
Exercise	0	0	0	5	358	98.62

[2] D. E. Warburton, C. W. Nicol, and S. S. Bredin, "Health benefits of physical activity: the evidence," *Canadian medical association journal*, vol. 174, no. 6, pp. 801–809, 2006.

[3] C. Mathers, G. Stevens, and M. Mascarenhas, *Global health risks: mortality and burden of disease attributable to selected major risks*. World Health Organization, 2009.

[4] J. W. Lockhart, T. Pulickal, and G. M. Weiss, "Applications of mobile activity recognition," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 1054–1058.

[5] M. R. Puyau, A. L. Adolph, F. A. Vohra, and N. F. Butte, "Validation and calibration of physical activity monitors in children," *Obesity research*, vol. 10, no. 3, pp. 150–157, 2002.

[6] A. Brajdic and R. Harle, "Walk detection and step counting on unconstrained smartphones," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 2013, pp. 225–234.

[7] A. Akahori, Y. Kishimoto, and K. Oguri, "Estimate activity for m-health using one three-axis accelerometer," in *Medical Devices and Biosensors, 2006. 3rd IEEE/EMBS International Summer School on*. IEEE, 2006, pp. 122–125.

[8] E. Munguia Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure," Ph.D. dissertation, Massachusetts Institute of Technology, 2008.

[9] D. Maguire and R. Frisby, "Comparison of feature classification algorithm for activity recognition based on accelerometer and heart rate data," in *9th. IT & T Conference*, 2009, p. 11.

[10] G. Salton, E. A. Fox, and H. Wu, "Extended boolean information retrieval," *Communications of the ACM*, vol. 26, no. 11, pp. 1022–1036, 1983.

[11] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information processing & management*, vol. 24, no. 5, pp. 513–523, 1988.

[12] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1. IEEE, 2005, pp. 604–610.

[13] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval*. ACM, 2007, pp. 197–206.

[14] H. Jiawei and M. Kamber, "Data mining: concepts and techniques," *San Francisco, CA, itd: Morgan Kaufmann*, vol. 5, 2001.