Using an Indoor Localization System for Activity Recognition

Andrea Aliperti, José Corcuera, Chiara Fruzzetti, Gianluca Marchini, Francesco Miliani, Simone Musetti, Andrea Primaverili, Riccardo Rocchi, Davide Ruisi, and Alessio Vecchio[0000−0001−6894−7338]✉

University of Pisa, Pisa, Italy
alessio.vecchio@unipi.it

Abstract. Recognizing the activity performed by users is important in many application domains, from e-health to home automation. This paper explores the use of a fine-grained indoor localization system, based on ultra-wideband, for activity recognition. The user is supposed to wear a number of active tags. The position of active tags is first determined with respect to the space where the user is moving, then some position-independent metrics are estimated and given as input to a previously trained system. Experimental results show that accuracy values as high as $\sim 95\%$ can be obtained when using a personalized model.

Keywords: Activity recognition · Wearable device · Ultra-wideband

1 Introduction and Related Work

Automatic recognition of user's activity is useful in numerous application domains. In e-health, activity recognition can be used to understand if the user is characterized by a sedentary lifestyle or to automatically assess his/her training sessions [1]. In home automation, services and applications can be customized according to user's activities [2]. In industrial environments, automatic recognition of activities can be used to implement smart control systems or to put advanced safety policies in place [3]. From a more general perspective, activity recognition is considered a key element of context-aware computing, as it provides relevant information about the user and the surrounding environment [4].

Given its relevance, research on activity recognition has been rather prolific during the last years, both from the point of view of adopted technologies and inference methods. In particular, a wide range of sensing mechanisms has been explored to collect information about the user, from 3D cameras to smart textile [5, 6]. However, the vast majority of proposed approaches relies on wearable devices equipped with inertial sensors like accelerometers and gyroscopes [7]. Sensors attached to users' limbs are used to collect information about their raw movements and then, at a higher level, to determine which activity is currently performed [8].

In a seminal paper, Bao and Intille used five biaxial accelerometers to recognize physical activities [9]. Devices were worn at the right wrist, right hip, left arm, left tight, and right ankle, whereas classification was carried out using the following methods: decision table, k-Nearest Neighbors, C4.5 decision tree, and Naive Bayes. The methods that provided the best results were decision trees and k-Nearest Neighbors.

In [10], a single triaxial accelerometer positioned on the belt was used to recognize some activities of daily living. Classification was carried out using two methods: Naive Bayes and k-Nearest Neighbors. A leave-one-person-out strategy was adopted for validation: two users were involved in the training phase, and five other users were recruited for the evaluation. The set of considered activities included: jump, run, walk, sit, and the transitions from sitting to standing, and from standing to kneel (and viceversa).

A comparison of different architectural solutions and the effects of parameters of operation in activity recognition is presented in [11]. Solutions based on both a single sensor and multiple sensors were evaluated. As far as classification methods are concerned, Neural Network achieved the best accuracy results. However, if also the training costs are considered, decision trees were identified as one the best methods for the reference scenario. Additional experiments showed that solutions based on multiple sensors are able to achieve better recognition accuracy, even when using light-weight algorithms. Also the effects of sampling rate on recognition accuracy were analyzed. Results showed that sampling rates greater than 20 Hz may provide limited benefits.

Fusion of data originating from multiple accelerometers was studied also in [12]. Data from two accelerometers significantly improves the accuracy of activity recognition with respect to the use of a single device. Using three or more sensors, on the contrary, seems to bring only limited benefits. When two accelerometers are used, ankle and wrist were identified as particularly favorable positions.

Given their popularity, smartphones and smartwatches have been frequently considered as suitable sensing devices for recognizing human activities. In addition, smartphones and smartwatches are characterized by reduced invasiveness and thus provide the opportunity to make the recognition system more acceptable for the end users [13, 14]. Smartphones, for instance, were used to distinguish falls from other activities of daily living [15]. A wrist-worn accelerometer proved to be extremely effective in detecting gait segments, especially when used in combination with a model tailored to the user [16]. Also the strength of the WiFi signal received by a smartphone was used as a means for activity recognition [17].

This paper contributes to existing literature by showing that an indoor localization system characterized by high resolution can be used for activity recognition. Differently from the vast majority of known methods, the proposed approach does not rely on inertial sensors. The accurate position of a set of devices worn by the user is first estimated and then used to extract some characterizing features. Experimental results show that, for the considered set of activities, average accuracy values as high as $\sim 95\%$ can be obtained. For some users the

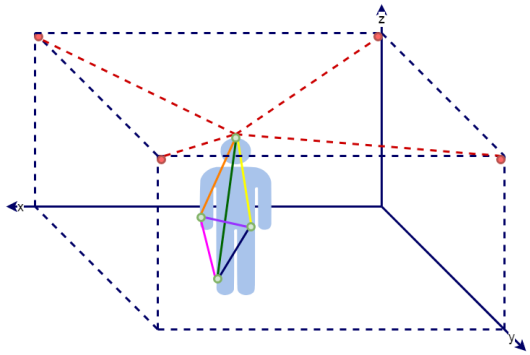

Fig. 1. The global reference system and the set of computed distances.

method is able to obtain perfect, or close to perfect, results. A prototypical implementation of the method based on IEEE 802.15.4-2011 Ultra-WideBand (UWB) is also described.

2 Method

The proposed activity recognition method is based on the idea of observing the user's movements by means of a fine-grained indoor localization system. A number of UWB-enabled devices with known position, called *anchors*, are placed in the environment. Anchors are able to determine the distances between themselves and *tags*, devices that are free to move in the considered space. Distances between tags and anchors are used to determine the tags' position, e.g. using a multi-lateration algorithm. We suppose that the user carries (wears) a number of tags, which could represent smart devices like fitness tracking wristbands, smartshoes, smartwaches, smartglasses or a simple smartphone carried in a pocket. The basic idea is that the position of these devices (tags) in the environment can be used to infer the activity currently performed by the user. In particular this is achieved by means of three steps: first, the position of tags in the environment[1] is used to compute a set of metrics that do not depend on the absolute position of the user; second, a set of features is extracted from these metrics; third, a previously trained machine learning algorithm is fed with feature values to determine which activity is currently performed. The three steps are detailed in the following.

2.1 From global positions to local metrics

Let us call A_1, A_2, ..., A_n the n anchors, and T_1, T_2, ..., T_m the m active tags worn by the user. Anchors periodically estimate the distances between themselves and

[1] We will also use the term *global* to indicate the environment-based reference system, as it is common to all the users.

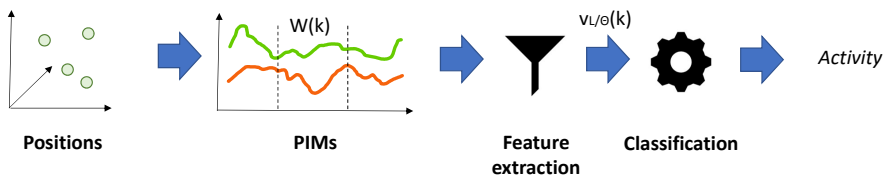

Fig. 2. Overview of the method

all the tags via UWB. The n distances between A_1, ..., A_n and the ith tag are used to derive the position of the latter in the global reference system. The position of the ith tag in the global reference system at time t and its components are indicated as $g_i(t) = \{g_i^X(t), g_i^Y(t), g_i^Z(t)\}$. Global positions are not directly used to infer the activity of the user, as they would make the activity recognition process dependent on the specific environment. To make the recognition process as general as possible, the position of tags in the global space is converted in a set of metrics that do not depend on the absolute coordinates of the devices. First, a set of differences between position vectors are calculated: $d_{ij}(t) = g_j(t) - g_i(t) = \{g_j^X(t) - g_i^X(t), g_j^Y(t) - g_i^Y(t), g_j^Z(t) - g_i^Z(t)\}$ with $i, j \in 1..m, j \neq i$. Then the following position independent metrics (PIMs) are calculated:

$$L_{ij}(t) = |d_{ij}(t)| \quad \text{with} \quad i, j \in 1..m, j > i$$

and

$$\Theta_{ijk} = \arccos\left(\frac{d_{ji}(t) \cdot d_{jk}(t)}{|d_{ji}(t)||d_{jk}(t)|}\right) \quad \text{with} \quad i, j, k \in 1..m, i \neq j, k \neq j, k > i$$

.

Hereafter, the two sets of PIMs will be indicated as L-PIMs and Θ-PIMs respectively. Figure 1 shows an exemplificative scenario with four anchors and four tags. The dashed lines represent the distances between one of the tags and the anchors placed in the environment (for the sake of image clarity only the distances between the head-mounted device and the anchors are shown). These distances are used by the indoor localization system to estimate the position of the tags in the global reference system. The four tags define a general tetrahedron, with the tags placed at the vertices. L-PIMs correspond to the length of the edges of the tetrahedron (the six continuous lines in the example), whereas Θ-PIMs are equal to the angles on the faces of the tetrahedron. The total number of angles is equal to twelve, but the space of freedoms, for tetrahedra, is five-dimensional [18]. Thus, the set of Θ-PIMs is not completely independent. Similar considerations may be made when the number of tags is smaller/greater than four.

2.2 Feature extraction

The set of PIMs is first preprocessed using a low-pass Butterworth filter. Processed signals are then segmented using a window with fixed duration R. Let

Fig. 3. Volunteer wearing four tags (head, right wrist, left pocket, right ankle)

us indicate the kth window concerning L_{ij} as $W_{ij}(k)$ (i.e. $L_{ij}(t)$ with $t \in [(k-1)R, kR)$). A set of functions is used to extract F characteristic indicators from $W_{ij}(k)$, for every k. In summary, a vector $v_L(k)$, the feature vector, is produced with period R. The same is done with Θ-PIMs, thus generating an analogous $v_\Theta(k)$ feature vector.

2.3 Inferring activities

The vector of feature values ($v_L(k)$, $v_\Theta(k)$, or both) is given as input to a previously trained machine learning classifier. For the training phase we explored two possible solutions, one based on a single model for all users and the other based on a personalized model. Both solutions have advantages and drawbacks: the first is characterized by a simpler training phase, as it does not require user-dependent customization; the second relies on a training phase that is user specific, but it is expected to provide better results.

An overview of the method is provided in Figure 2.

3 Prototype and Data Collection

A prototype of the system has been implemented using a Decawave MDEK1001 kit [19]. The kit is composed by a set of Decawave DWM1001-DEV boards, and a real-time localization application. Each DWM1001-DEV board is equipped with a DW1000 IEEE 802.15.4-2011 UWB transceiver and a Nordic nRF52832 BLE microcontroller (based on ARM Cortex M4). The Bluetooth connection was used only to define a network and to set the parameters of operation. Four boards were used as anchors and attached to the walls at the corners of a square area

User ID	Age (yr)	Weight (kg)	Height (m)	Gender (M/F)
1	24	79	1.72	M
2	24	55	1.66	F
3	25	76	1.87	M
4	33	74	1.72	M
5	24	80	1.80	M
6	25	80	1.80	M
7	25	77	1.82	M
8	26	45	1.43	F
9	24	80	1.82	M
10	23	66	1.75	M

Table 1. Ten users participated in the data collection phase

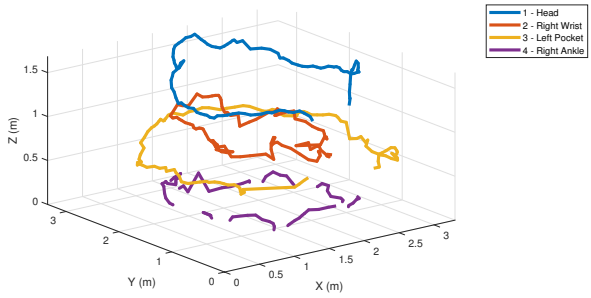

Fig. 4. 3D trajectories of the four worn devices when the volunteer is walking in circle

of 3.60 m x 3.60 m (corresponding to a $\sim$ 13 m^2 room). Anchors were placed 2 meters above the ground. Four other boards were used as tags and attached at the head, right wrist, left pocket, and right ankle of the user. Figure 3 shows a volunteer wearing the prototype. Another board operating as a listener was attached to a common PC, where the position of tags was logged. The position of tags was estimated with 10 Hz frequency, the maximum allowed by the current version of the kit.

Ten volunteers participated in the data collection phase. Their main characteristics are listed in Table 1. Each volunteer was asked to perform the following activities: standing, sitting, walking in circle, laying on the floor, and crouching. Each activity was performed for 60 s. The activities of all volunteers were video recorded. This was done to simplify the subsequent manual labeling of acquired traces and to document the whole process in case of unexpected results. Figure 4 shows a 3D representation of the walking activity of one of the users (approximately 12 seconds, corresponding to an almost-complete circle). The trajectory of the ankle-worn device is not always visible because of positioning errors producing negative values on the Z axis.

All traces were segmented using 3 s windows ($R = 3$ s). Traces were then manually annotated assigning one of the following labels to each window: *stand-*

	Mean	Max-Min	Min	AAV	Std.Dev.	RMS	MCR	Max	MAD
L_{12}	✓		✓			✓			
L_{13}	✓	✓	✓	✓		✓		✓	
L_{14}	✓		✓	✓		✓		✓	
L_{23}	✓		✓			✓			
L_{24}	✓		✓	✓		✓		✓	✓
L_{34}	✓	✓	✓	✓	✓	✓			✓

Table 2. Selected features

ing, *sitting*, *walking*, *lying*, *crouching*, *transition*, and *invalid*. The first five labels were used as ground truth during the training and evaluation phases. The *transition* label was used to identify the time windows containing multiple activities (e.g. when the user stops walking and starts lying on the floor). The last label was used to mark a limited number of segments containing data to be discarded (in few occasions, one of the tags suffered from short disconnections).

Labeled traces are publicly available at the following address:
`URL will be provided in case of publication`

4 Results

The following set of statistical and signal processing functions was considered for the feature extraction phase: mean, max-min, min, average absolute variation (AAV), standard deviation, root mean square (RMS), mean crossing rate (MCR), max, mean absolute deviation (MAD). AAV is a function successfully used in similar contexts [20], and it is defined as follows:

$$AAV = \frac{1}{N}\sum_{i=1}^{N}|a_{i+1} - a_i|$$

where a_i is the ith observed value and N is the number of values in the considered window.

First, we evaluated the performance of the method using only the set of L-PIMs. With $m = 4$, a vector v_L with 54 features was produced every R seconds (the size of v_L is equal to the number of functions multiplied by the number of different L-PIMs). The number of features was then reduced to 30 using the RELIEFF method [21]. The final set of features is shown in Table 2.

Transitions between different activities are not considered in this study, as we are interested in understanding which is the performance of the method in steady conditions. However, an approach like the one proposed in [22] can be possibly incorporated.

The reduced set of features was given as input to a number of machine learning methods (those available in the MATLAB classification learner toolbox). Evaluation was carried out according to ten-fold cross validation. The dataset was divided in ten disjoint subsets. Nine subsets were used to train the machine

Classification method	Accuracy (%)
ENS	93.0
SVM	92.6
KNN	92.0

Table 3. Results obtained by some well-known classifiers

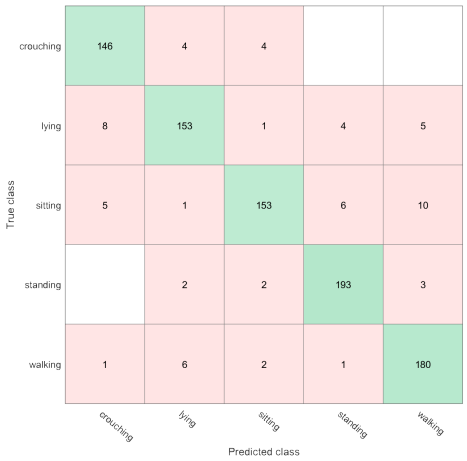

Fig. 5. Confusion matrix

learning method, and the left out subset was used to evaluate the method on previously unseen data. The process was repeated ten times using all the different subsets for the evaluation phase and then averaging the results. Table 3 shows the obtained accuracy values for the top three performing methods: Ensemble Subspace KNN (ENS), Support Vector Machine (SVM), k-Nearest Neighbors (KNN). The three classifiers provided very close results with 92-93% accuracy. The confusion matrix for the set of considered activities is shown in Figure 5.

Such results were obtained training the system using the data of all users. This approach corresponds to building a global model of users' activities. We then evaluated the possible benefits introduced by a personalized model. To this aim, we performed the feature selection, training, and evaluation phases again using the data of a single user at a time. Also in this case ten-fold cross validation was used to avoid overfitting. Results are reported in Table 4. When using a personalized model, accuracy values are generally better then the ones obtained by the global model. On average, the personalized approach is able to obtain accuracy values that are close to $\sim 95\%$. For some users, accuracy values are equal, or very close to, 100%. In summary, the use of a personalized model has a positive impact on the proposed method.

We then repeated the same analysis using the set of Θ-PIMS. In particular, six different Θ values were computed and given as input to the same classification algorithms (the number of degrees of freedom, as mentioned, is equal to five, but we decided to include one more value to mitigate the effects of measurement

	Accuracy (%)		
User ID	ENS	SVM	KNN
1	96.8	95.7	95.7
2	97.9	97.9	97.9
3	96.7	97.8	97.8
4	100.0	97.7	98.9
5	98.8	97.7	98.8
6	86.2	87.5	83.7
7	96.7	94.4	95.6
8	95.7	96.8	96.8
9	93.7	95.0	95.0
10	85.6	87.8	84.4
Mean	94.8	94.8	94.4

Table 4. Accuracy for the different users when using a personalized model

errors). The number of features was again limited to 30 using RELIEFF. The set of selected features is similar to the one produced for L-PIMs, with mean, min, RMS, max, and AAV the most popular functions (selected 6, 6, 6, 5, and 3 times respectively). Also in this case we considered both the approach based on the definition of a global model and the one based on a personalized model. With the global model, the following accuracy values were obtained by ENS, SVM, and KNN respectively: 91.3%, 91.2%, and 91.9%. With the personalized model, the obtained accuracy values were equal to 94.4%, 93.9%, and 94.0% Results obtained with Θ-PIMs are thus approximately equal to the ones obtained with L-PIMs. Finally, the same analysis was carried out using both L- and Θ-PIMs (but still limiting the set of features to the best 30). With the global model, the following accuracy values were obtained by ENS, SVM, and KNN respectively: 92.4%, 93.4%, and 93.8%. With the personalized model, the obtained accuracy values were equal to 94.6%, 95.5%, and 94.9%, thus achieving slightly better final results.

5 Conclusion

Recent advancements in communication technologies, e.g. IEEE 802.15.4-2011 UWB, not only make fine-grained localization of devices and people possible, but also provide the opportunity to devise new applications or to improve existing ones. This paper shows that an indoor localization system can be successfully used for automatic activity recognition. To be as general as possible, the proposed approach relies on a set of metrics and features that are not dependent on the position of the user in the environment. Experimental results show that excellent accuracy values can be achieved, especially when using a model tailored to the user. Obviously, it is also possible to combine PIMs with the user's position in the environment (already provided by the indoor localization system), to further improve the recognition process. For instance, if the user is in the kitchen the set

of possible activities could be expanded/reduced with respect to the one used in this paper (e.g. adding cooking and removing lying).

Acknowledgment

This work was funded in part by University of Pisa, grant PRA_2017_37 "IoT and Big Data".

.

References

1. M. Ermes, J. Prkk, J. Mntyjrvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 20–26, Jan 2008.
2. A. Benmansour, A. Bouchachia, and M. Feham, "Multioccupant activity recognition in pervasive smart home environments," *ACM Comput. Surv.*, vol. 48, no. 3, pp. 34:1–34:36, Dec. 2015. [Online]. Available: http://doi.acm.org/10.1145/2835372
3. J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner, "Activity recognition of assembly tasks using body-worn microphones and accelerometers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1553–1567, Oct 2006.
4. N. Davies, D. P. Siewiorek, and R. Sukthankar, "Activity-based computing," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 20–21, April 2008.
5. J. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review," *Pattern Recognition Letters*, vol. 48, pp. 70 – 80, 2014, celebrating the life and work of Maria Petrou. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865514001299
6. J. Cheng, O. Amft, G. Bahle, and P. Lukowicz, "Designing sensitive wearable capacitive sensors for activity recognition," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3935–3947, Oct 2013.
7. A. Bulling, U. Blanke, and B. Schiele, "A tutorial on human activity recognition using body-worn inertial sensors," *ACM Comput. Surv.*, vol. 46, no. 3, pp. 33:1–33:33, Jan. 2014. [Online]. Available: http://doi.acm.org/10.1145/2499621
8. O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors." *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013.
9. L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*, A. Ferscha and F. Mattern, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 1–17.
10. P. Gupta and T. Dallas, "Feature selection and activity recognition system using a single triaxial accelerometer," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1780–1786, June 2014.
11. L. Gao, A. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Medical Engineering & Physics*, vol. 36, no. 6, pp. 779 – 785, June 2014.

12. A. K. Chowdhury, D. Tjondronegoro, V. Chandran, and S. G. Trost, "Physical activity recognition using posterior-adapted class-based fusion of multiaccelerometer data," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 678–685, May 2018.
13. J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *SIGKDD Explor. Newsl.*, vol. 12, no. 2, pp. 74–82, Mar. 2011. [Online]. Available: http://doi.acm.org/10.1145/1964897.1964918
14. T. Brezmes, J.-L. Gorricho, and J. Cotrina, "Activity recognition from accelerometer data on a mobile phone," in *International Work-Conference on Artificial Neural Networks*. Springer, 2009, pp. 796–799.
15. S. Abbate, M. Avvenuti, F. Bonatesta, G. Cola, P. Corsini, and A. Vecchio, "A smartphone-based fall detection system," *Pervasive Mob. Comput.*, vol. 8, no. 6, pp. 883–899, Dec. 2012. [Online]. Available: http://dx.doi.org/10.1016/j.pmcj.2012.08.003
16. G. Cola, M. Avvenuti, F. Musso, and A. Vecchio, "Personalized gait detection using a wrist-worn accelerometer," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, May 2017, pp. 173–177.
17. S. Sigg, U. Blanke, and G. Trster, "The telepathic phone: Frictionless activity recognition from wifi-rssi," in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, March 2014, pp. 148–155.
18. R. Andr and F. P. W., "Is there a most chiral tetrahedron?" *Chemistry A European Journal*, vol. 10, no. 24, pp. 6575–6580. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.200400869
19. "Decawave," http://www.decawave.com, accessed: 2018-05-15.
20. G. Cola, M. Avvenuti, A. Vecchio, G. Z. Yang, and B. Lo, "An unsupervised approach for gait-based authentication," in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, June 2015, pp. 1–6.
21. I. Kononenko, E. Šimec, and M. Robnik-Šikonja, "Overcoming the myopia of inductive learning algorithms with relieff," *Applied Intelligence*, vol. 7, no. 1, pp. 39–55, 1997.
22. J.-L. Reyes-Ortiz, L. Oneto, A. Sam, X. Parra, and D. Anguita, "Transition-aware human activity recognition using smartphones," *Neurocomputing*, vol. 171, pp. 754 – 767, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231215010930