

AKADÉMIAI KIADÓ

Convolutional neural network-based activity monitoring for indoor localization

László Árvai* 

Hatvany József Doctoral School of Information Science, University of Miskolc, H-3515 Egyetemváros, Miskolc, Hungary

Received: December 31, 2020 • Revised manuscript received: April 6, 2021 • Accepted: April 8, 2021

Published online: May 31, 2021

Pollack Periodica •
An International Journal
for Engineering and
Information Sciences

16 (2021) 3, 7–12

DOI:

[10.1556/606.2021.00370](https://doi.org/10.1556/606.2021.00370)

© 2021 The Author(s)

ORIGINAL RESEARCH
PAPER



ABSTRACT

Location specific services are widely used in outdoor environment and their indoor counterpart is gaining more popularity as well. There is no standardized technology exists for indoor localization, usually smart phone is used as a localization platform and the field strength of an existing radio frequency infrastructure is used as the location specific information. Smart devices are also equipped with several sensors capable of capturing the motion data of the device. Detecting the walking step, turn, stairs motion type can refine the indoor position using digital indoor map as a reference. The real-time recognition of the motion type is possible with a precisely constructed and trained convolutional neural network and therefore it can improve the stability of the localization.

KEYWORDS

indoor-localization, convolutional neural network, human activity monitoring

1. INTRODUCTION

The indoor localization is an extensively researched area nowadays. For outdoor localization there is a well-known and widely used Global Positioning System (GPS) technology, but there is no standardized and commonly used method exists for location determination in indoor environment. Therefore, in the last few years several research activity and proposal emerged but there still no common solution is available.

The proposed solutions can be divided into two categories. The first category requires a special infrastructure, built for the localization. This method provides very high accuracy; even sub cm accuracy can be achieved in the expense of specially built infrastructure. Signal transmitters or special markers needs to be installed and maintained.

The other group of the indoor localization methods are using some existing infrastructure for the localization which has an advantage of the low installation and maintenance effort but usually this comes with lower accuracy, most of the cases it is not better than few meters. For many applications this precision is appropriate, for example navigating in a shopping mall or big railway station. The other advantage of these systems is their sensing or navigation device is a very often only a smartphone, which is very popular, cheap and widespread. For this solution only specific software must be installed onto the mobile phone, which is capable of performing all calculation for the location determination.

Developing and indoor navigation software is not an easy task since there is no localization specific infrastructure and usually the existing infrastructure is not directly applicable for the location determination, different data processing and filtering algorithm needs to be implemented in order to provide location information with acceptable precision.

*Corresponding author.

E-mail: laszlo.arvai@gmail.com

2. INDOOR LOCALIZATION

The indoor localization algorithm must be able to run on a regular smart phone, external fileservers are used only for data storage. After retrieving and locally caching the appropriate navigational data the server connection is not required anymore. It means the calculation power and measurement data source applicable for the indoor localization are only the processing power and the built-in sensors of a smart phone. The latest, usually are the Wi-Fi communication module, the inertial sensors (three axis acceleration, angular velocity, magnetic field) and air pressure (barometer) sensor.

2.1. Related work

The simplest solution for indoor localization is analyzing the user's motion and predicting the trajectory based on this observation. The user's motion can be observed from outside, using cameras and other optical sensors [1]. This solution provides the most accurate result, but it is not an option for a mobile phone based indoor navigation system. A simple example scenario for motion analysis is finding a car in a parking lot, described in [2]. It proposes to use the inertial sensor data with a Kalman filter to estimate the user's trajectory and for request it can be played back in reverse order to guide the user back to the originated place. This is a typical Pedestrian Dead Reckoning (PDR) application, which is well known for integrating the error over time, so it might work without significant error filtering effort only in short distances.

Combining the PDR with other location refining method can improve the reliability as it shown in [3]. This paper proposes using a preexisting digital floor-plan to check the users walking and turning activities against this map and refine the position with the information where the specific motion type is possible. It also proposes a simple heuristics method for the step detection. The proposed method only detects the steps and it is not able to classify it. The classification of the motion might provide additional information about the current navigation scenario.

The motion classification problem is basically a time series classification problem, since the mobile phone sensors are providing time series data about the acceleration and angular velocity of the user. There are many possible ways for this kind of classification. One of proposed solution [4] uses fuzzy classifier for human motion analysis and hidden Markov model for the context interpretation. Even this solution has an advantage to be able to work without any external infrastructure; the accuracy - especially of the human activity recognition - is relatively low, which might render low positioning accuracy.

The next logical improvement is using some location specific information beside the PDR method to increase the long-term reliability of the algorithm. The simplest location specific information is detecting the close proximity to a known position. These predefined positions can be marked by NFC tags and the proximity of the mobile phone can be

detected by the built-in NFC reader as it is show in [5]. It also uses neural network for cleaning the acceleration sensor data, but not for step classification and detection. The most popular and still special infrastructure-less solution is using Wi-Fi or Earth magnetic field strength. Since these are location specific and using for example a particle filter, they can be easily combined with a PDR. The [6, 7] introduces a solution, using particle filter and mainly Wi-Fi radio map for the localization. The motion analysis however very simple, they are using zero velocity update method, which provides reliable step detection but cannot determine the motion type.

The motion type determination or Human Activity Detection (HAR) is a very popular research topic recently. With the availability of smart watches and mobile devices the activity monitoring can be easily implemented and it is a useful data source for health or fitness applications. For this purpose, many classification algorithms are proposed, like Support Vector Machines (SVM) [8] or a low computational version of SVM [9].

Using Artificial Intelligence (AI), Machine Learning (ML) methods for time series analysis are gaining more attention in many fields [10]. In the case of motion analysis, the most accurate method is using Convolutional Neural Network (CNN) as it is shown in [11] and [12]. They compare it with SVM and Random Forest and prove superiority of CNN in terms of recognition accuracy. The [13] also proposes a simplification by instead of using three-dimensional acceleration vector for the classification; they are only using the vector length, reducing the problem into 1D time series classification. The advantage of this method is removing the mobile phone directional dependency from the classification but in the same time it removes useful features from the classification.

The proposed methods are not providing real-time motion recognition. They are using 2-3 seconds long time window for the classification, but for fast walking this time frame might contain more than one steps, which means it has a chance to lose step event information.

The closest solution to the required real-time analysis is proposed by [14]. This method gives real-time operation and very high recognition accuracy. The only problem is the dataset for training of this network does not contain all of the required motion type, it lacking by escalator, elevator movement and turns.

2.2. Proposed system

Considering the aforementioned constrains the most applicable localization algorithm is based on Wi-Fi field strength sensing, combined with user's motion analysis. The motion analysis is done by CNN and the localization activity is controlled by a particle filter.

For the position estimation using field strength value requires a previously measured field strength map or fingerprint map of the area. This activity must be done in advance, using a mobile phone and appropriate software for field strength measurement.

Beside the field strength map, the other requirement for the algorithm is the existence of the indoor floor-plan (geometry of the indoor scenery) in digital form with the appropriate metadata (properties of indoor elements) for example the type of the element like doors, stairs, elevators and other relevant properties. This digital map is created using a toolset provided by the Open Street Map (OSM) community. The OSM is an open-source digital map system originally created for outdoor map handling but it also has a useful addition for indoor map creation [15]. It has a tool called Java Open Street Map (JOSM) editor [16] for creating the vector based indoor maps and defining all metadata for each element on the map.

Once the digital map is created and all field strength measurements are done the resulting database, so called localization infrastructure can be downloaded to the mobile phone and the actual indoor localization algorithm can be executed.

The localization algorithm has two phases as it can be seen on Fig. 1. The predictor phase is based on the user's motion analysis and estimates the user's position using the previous position updated with the sensed motion vector. The motion vector checked against the geometric constrains, for example walls are blocking the motion while doors are not.

The corrector phase is when the newly estimated position is compared with the Wi-Fi field strength map. It uses field strength information from the infrastructure database and uses k-nearest neighbor search for finding the closest matching fields strength vectors.

The two phases controlled by a particle filter [17]. The particle filter operates with several position estimations (particles) and updates them using the predictor and corrector in iterative way. In each iteration step the group of particles is resampled, only the most probable estimations (based on the corrector stage measurement) will be used in the next step. The final location estimation can be obtained by the weighted average of all particles where the weights are the probabilities assigned to each particle.

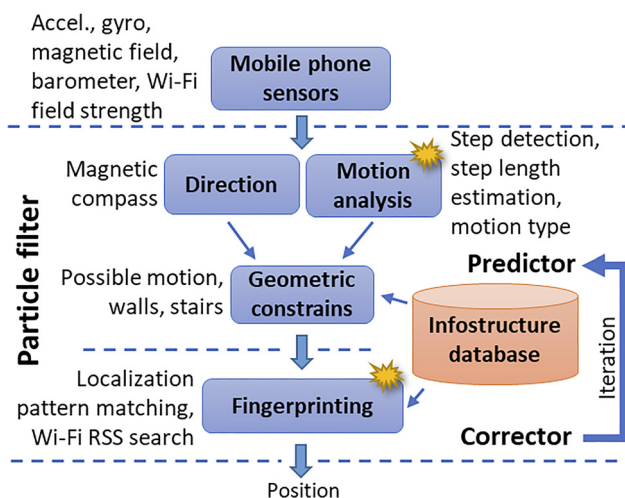


Fig. 1. Indoor localization algorithm

The Wi-Fi Access Point (AP) and field strength value filtering are also important part of the algorithm but since it is explained in other paper [18] it is not repeated here.

The particle filter method can be supported at two places by convolution neural network-based activity detection; these modules are marked by stairs in Fig. 1. The motion analysis module can use the detection of a walking step event and combining it with the step length estimation and direction information and can provide motion trajectory estimation. For this purpose, reliable and real-time step detection must be provided by the neural network.

The second module, where the activity detection can be used is the corrector phase. The probability (weight) determination of estimated location (particle) can use the current activity information. For example, if the currently recognized activity is walking upward on a staircase, the particles located on staircase gets higher probabilities. The example positions where activity recognition can be used to refine the location estimation are shown in Fig. 2. On the image a hypothetical route is shown from the staircase to the elevator. The circles (orange) are the markers of the turning points. Recognizing the significant turn in the user's motion can reinforce this position estimation. The rectangles (red) are the markers of the motions containing vertical components. The recognition of usage of elevator, or walking on staircase can reinforce position estimation on these positions.

For this purpose - same as the step detection - an almost real-time activity recognition is required and obviously the digital map with appropriate metadata is also required.

The conclusion of the integration of convolutional neural network into a particle filter based indoor localization algorithm is the strongly expected real-time operation of the neural network, capable of recognizing each of the walking steps and determining the motion type. At least the level walking, significant turns, stairs up and down walking and elevator usage must be recognized. The sensory data for the

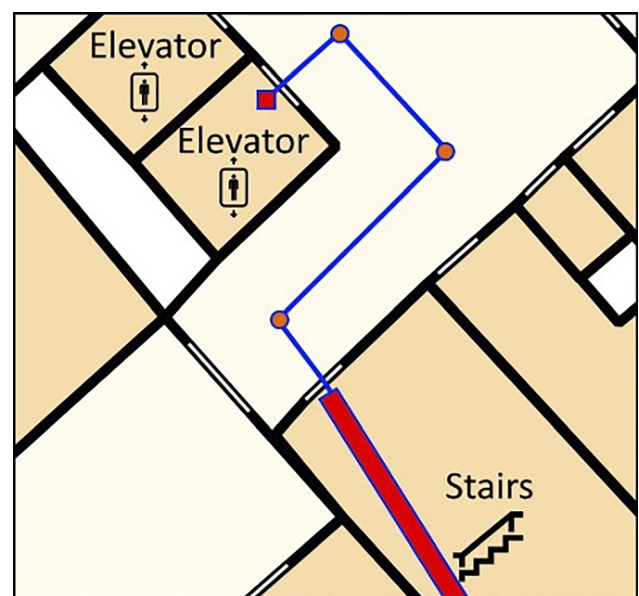


Fig. 2. Activity recognition positions

recognition activity is mainly the acceleration sensor, for turns the angular velocity sensor and for the recognition of vertical motion the pressure sensor might be used.

In order to construct a neural network, the appropriate data must be collected and preprocessed for training the network.

3. DATA COLLECTION AND PREPROCESSING

The data collection was done using a mobile phone as a collection device and having several volunteers for participating in the procedure. A data collection application for mobile phone has been developed, which collects the sensor data with the highest possible sample rate and stores it in the phone's non-volatile memory. The phone was placed in two positions: trouser pocket and in hand (horizontally, same as the usual position for navigation application when the user can see screen of the phone). In the measurement 30 participants were involved, 18 males and 12 females, between age 21–50 years with different body height.

The measurement protocol contained different kind of motions. The participants had to walk straight and level (20 steps) with three different walking speed, make turns, walk stairs up and down. They also had to use nine different type of elevator for measuring the motion data while travelling in elevator both ascending and descending directions.

The collected data consists other, non-walking related motion types as a negative reference. This motion includes travelling by a car or bicycle, regular office activities (sitting, typing, reading or even eating).

3.1. Data filtering

Since the measured values of inertial sensors is covered by noise [19] the collected data is filtered before further processing it. The same filtering method is applied to the learn dataset and the real-time data for the actual activity detection. The data from the acceleration sensor is filtered in two steps. The first step is provided by the Android Operation System (OS), since besides reading the raw three axis accelerometer data, there is a function for getting the linear acceleration value. The linear acceleration is composed from raw acceleration data by subtracting the current gravity acceleration vector from it. It means this value only contains the acceleration caused by the user's movement and constant gravity value must not be considered and will not reduce the neural network ability to recognize activity in variable gravity conditions.

The second stage of the filtering is mainly the sample rate reduction. It is done by applying an up-sampling, cubic interpolation to increase the sample rate to constant 1000 Hz from whatever the mobile phone has (usually between 100 Hz–500 Hz). Then the signal is filtered using a 7th order Butterworth filter with 8 Hz cut-off frequency and finally the signal is down-sampled to 50 Hz, which is more than enough for activity detection. With this data flow the sample rate will be fixed for any mobile phone and also high

frequency noise of the acceleration sensor is reduced. The gyro sensor data flow is similar, but the main difference is the applied filter type is 5th order Butterworth high-pass filter with 1 Hz corner frequency, since the gyro has low frequency noise (drift).

The filtered sensor signals then must be labeled for using as learning dataset for the neural network.

3.2. Dataset labeling

The recognizable features of the activity must be tagged in the collected data. For tagging an application had been developed, which can execute the data filtering explained earlier and provides an easy-to-use toolset for the user to add activity information to the data.

All different dataset (walking, staircase, etc.) must be tagged and must be stored for the learning dataset. Then the neural network can be constructed and trained.

4. ACTIVITY CLASSIFICATION

The activity classification is basically a time series classification where the time series data is provided by the inertial sensor of the mobile phone. The convolutional neural networks are well suited for this task, as explained in details in [20]. The structure of the CNN has been used is very similar to [14] with the addition of gyro data, mostly for turn recognition.

The data processing is shown in Fig. 3 and starts with the sensors signals and with filtering, down-sampling. The resulting vector (3 acceleration and 3 gyro) data is then transferred to the convolution layer in 16 sample wide window. The layer has 196 feature maps followed by a max-pooling layer with size 1×4 to reduce number of features to the quarter of maps. The output of the max-pooling layer is coupled to a fully connected layer with 1024 neurons. The final layer is a softmax layer to provide probability of the five possible activities.

For training the network the standard stochastic gradient descent with back-propagation algorithm, combined with the Adam [21] optimization for network parameter optimization has been used. The model is trained to minimize cross-entropy loss function using the well-known Tensorflow application framework.

For the network verification the 30% of the dataset has been kept as a test dataset. The results of the recognition ability of the activities in the test dataset are shown in Table 1.

The step and staircase recognition has acceptable accuracy; however, the elevator recognition is relatively low. The cause of this problem is the relatively small number of elevator data. This part of the dataset needs to be expanded in the future and escalator data also needs to be added. The "Other" in the table shows the false identification of other activities as moving activity. This value is also planned to be reduced by expanding the training dataset.



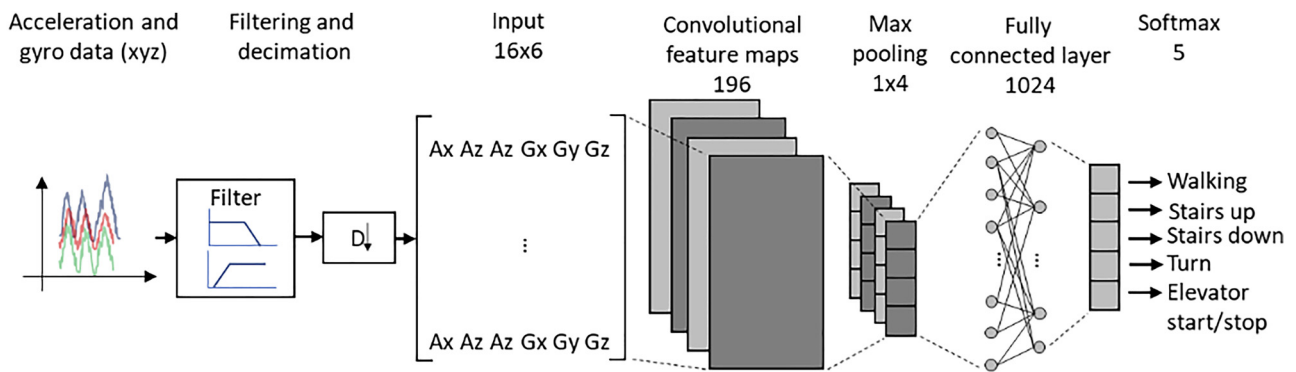


Fig. 3. Proposed convolutional neural network structure

Table 1. The results

Activity type	Accuracy %
Walk	91.4
Stairs Up	89.6
Stairs Down	88.9
Turn	86.3
Elevator start/stop	75.2
Other	8.2

One final activity is left, to port the whole trained CNN to TensorFlow Light in order to be able to run on mobile phone and to be integrated into the final mobile phone indoor localization application.

5. CONCLUSION

The paper presented the method how the particle filter based indoor localization algorithm can be improved by using convolutional neural network for human activity detection. The activity detection can be used in the prediction of the user motion by recognizing the steps of the user. It also can be used in the corrector phase, recognizing the motion type and using this information as a localization context data. The exact position can be refined by the possible motion type, like walking on staircase or using an elevator. This information can be successfully extracted from the sensor data of a mobile phone in real-time and can be integrated into indoor localization algorithm. The construction of a neural network for this purpose consists of several activities; the training dataset must be gathered by measuring many possible motion scenarios and person types. This dataset must be filtered and labeled for the training and the appropriate neural network structure must be constructed and trained. After this work a neural network with an acceptable accuracy can be obtained.

In the future, more optimization effort is required since the computing complexity of this method is not considered in this work; however it will significantly affect the power consumption of the mobile phone, which must be as minimal as possible. Also, the accuracy of the algorithm in some

special cases like elevator must be improved and other motion types like escalator should be added.

Finally, it can be concluded that the convolutional neural network-based activity recognition can help to increase the mobile phone based personal indoor localization system's performance.

REFERENCES

- [1] P. Müller and Á. Schiffer, "Human gait cycle analysis using Kinect V2 sensor," *Pollack Period.*, vol. 15, no. 3, pp. 3–14, 2020.
- [2] H. Fathy, R. Elbasiony, and M. T. Faheem, "Finding a parked car location in a multi-storey building without GPS service," *Int. J. Interactive Mobile Tech.*, vol. 14, no. 10, pp. 7–21, 2020.
- [3] J. Á. B. Link, P. Smith, N. Viol, and K. Wehrle, "FootPath: Accurate map-based indoor navigation using smartphones," in *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation*, Guimaraes, Portugal, Sept. 21–23, 2011, 2011, pp. 1–8.
- [4] A. Török, A. Nagy, L. Kováts, and P. Pach, "DREAR - Towards infrastructure-free indoor localization via dead-reckoning enhanced with activity recognition," in *Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, Oxford, England, Sept. 10–12, 2014, 2011, pp. 106–111.
- [5] T. Tajti, G. Geda, B. Tamás, and Gy. Vad, "Indoor localization using NFC and mobile sensor data corrected using neural net," in *Proceedings of the 9th International Conference on Applied Informatics*, Eger, Hungary, Jan. 29–Feb. 1, 2014, 2014, vol. 2, pp. 163–169.
- [6] Z. N. Z. Hongbo and F. W. W. Zulin, "A novel particle filter approach for indoor positioning by fusing WiFi and inertial sensors," *Chin. J. Aeronautics*, vol. 28, no. 6, pp. 1725–1734, 2015.
- [7] C. Real Ehrlich and J. Blankenbach, "Indoor localization for pedestrians with real-time capability using multi-sensor smartphones," *Geo-spatial Inf. Sci.*, vol. 22, no. 2, pp. 73–88, 2019.
- [8] Z. Y. He and L. W. Jin, "Activity recognition from acceleration data using AR model representation and SVM," in *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, Kunming, China, July 12–15, 2008, 2008, pp. 2245–2250.
- [9] D. Anguita, A. Ghio, L. Oneto, X. Parra, and J. L. Reyes-Ortiz, "Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine," in *Ambient Assisted*

- Living and Home Care*. Lecture Notes in Computer Science, vol. 7657, J. Bravo, R. Hervás, and M. Rodríguez, Eds., Berlin, Heidelberg: Springer, 2012, pp. 216–223.
- [10] I. Bolkeny and V. Fuvesi, “AI based predictive detection system,” *Pollack Period.*, vol. 13, no. 2, pp. 137–146, 2018.
- [11] B. Zhou, J. Yang, and Q. Li, “Smartphone-based activity recognition for indoor localization Using a convolutional neural network,” *Sensors*, vol. 19, no. 3, pp. 1–15, 2019.
- [12] C. A. Ronao and S.-B. Cho, “Deep convolutional neural networks for human activity recognition with smartphone sensors,” in *22nd International Conference on Neural Information Processing*, Istanbul, Turkey, Nov. 9–12, 2015, 2016, pp. 235–244.
- [13] S. M. Lee, S. M. Yoon, and H. Cho, “Human activity recognition from accelerometer data using convolutional neural network,” in *IEEE International Conference on Big Data and Smart Computing (BigComp)*, Jeju, South Korea, Feb. 13–16, 2017, 2017, pp. 131–134.
- [14] A. Ignatov, “Real-time human activity recognition from accelerometer data using convolutional neural networks,” *Appl. Soft Comput.*, vol. 62, pp. 915–922, 2018.
- [15] *OSM Simple Indoor Tagging*, Open Street Map Community, 2020. [Online]. Available: https://wiki.openstreetmap.org/wiki/Simple_Indoor_Tagging. Accessed: Dec. 1, 2020.
- [16] *JOSM is an extensible editor for OpenStreetMap (OSM) for Java 8+*, Open Street Map Community, 2020. [Online]. Available: <https://josm.openstreetmap.de/>. Accessed: Dec. 1, 2020.
- [17] S. Thrun, W. Burgard and D. Fox, *Probabilistic Robotics*. The MIT Press, 2005.
- [18] L. Árvai and Sz. Homolya, “Filtering and fingerprint matching methods for Wi-Fi radio map based indoor localization,” in *Proceedings of the Carpathian Control Conference*, Krakow-Wieliczka, Poland, May 26–29, 2019, 2019, pp. 704–709.
- [19] O. J. Woodman, “An introduction to inertial navigation,” *Technical Report*, no. 696, University of Cambridge, Computer Laboratory, 2007, pp. 1–37.
- [20] B. Zhao, H. Lu, S. Chen, J. Liu, and D. Wu “Convolutional neural networks for time series classification,” *J. Syst. Eng. Electronics*, vol. 28, no. 1, pp. 162–169, 2017.
- [21] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

