People identification using Kinect sensor

Adnan Ramakić[1], Amel Toroman[2]

[1] University of Bihac, Rectorate
adnan.ramakic@unbi.ba

[2] University of Bihac, Technical faculty
amel.toroman@unbi.ba

Abstract. In this work, we address a problem of people identification. People identification is an important feature in various application like using in banks, airports, border crossings etc. For purpose of people identification today are used different methods such as face recognition, fingerprint, scanning of eye retina, voice recognition etc. The most of these methods require interaction with people while one method, people gait recognition, can be proceeding even without awareness of people who is in process of identification. Because of that, people gait recognition is interesting field in identification process and biometrical techniques. Our approach for this imply using Kinect sensor from Microsoft and Matlab high level technical computing language. For image classification we use bag of features or bag of words. Process consists of extracting regions, compute descriptors, find clusters, and compute distance matrix and using SVM (Support Vector Machine) for Classification. Dataset which is used in this process is also created with Kinect sensor.

Keywords: People identification, gait recognition, Kinect sensor, bag of features, SVM

1 Introduction

People identification is very important task in many areas of human life. This task is provided in banks, airports, border crossing, in companies as security check etc. There are different methods for completing this task like face recognition, fingerprint, analyzing of eye (retina, iris), voice recognition etc. These methods are realized on some different ways but ultimately they do a job. The most of these methods require some kind of interaction with a person which passing a process of identification. One of methods which does not require interaction with persons is gait recognition. In this work we present a method for people identification in gait, but we are not analyzing gait features (like step length, time between steps etc.). Instead we use whole pictures in dataset which contains persons in different gait position (these positions are positions captured while persons walk in one direction) and based on that, after training phase, prediction is done in real time. For reasons that we analyze pictures with persons in gait position state of the art is focused on works with this topic. Many works are done with gait

recognition and basically are model based or appearance based. Sivapalan et al. [1] extend a Gait energy images concept (GEI) with 3D and represent Gait energy volume (GEV). Sivapalan et al. [2] also present Backfilled GEI (BGEI), a new capture-modality independent feature. Hofmann, Bachmann i Rigoll [3] expand in their approach GEI concept with depth information and present Depth Gradient Histogram Energy Image (DGHEI). Borràs, Lapedriza i Igual [4] represent a DGait database gained with depth camera and extract 2D and 3D gait features based on shape descriptors and compare the performance of these features for gender identification using Kernel SVM (Support vector machine). Lu, Wang i Moulin [5] are in their work present Sparse reconstruction based metric learning method (SRML). Chattopadhyay et al. [6] are explore the applicability of Kinect RGB-D streams in recognizing gait patterns of individuals. They register the depth and RGB frames from sensor to obtain smooth silhouette shape along with depth information where partial volume reconstruction of the frontal surface of each silhouette is done. In they work proposed gait feature is called Pose Depth Volume (PDV). Arora i Srivastava [7] proposed spatial-temporal-based method for human gait recognition called Gait Gaussian Image (GGI). Preis et al. [8] present the approach based on Microsoft Kinect where they evaluate a number of body features together with step length and speed, while in [9] authors integrate depth information in a silhouette-based gait recognition scheme to produce hybrid 2D-3D frontal gait recognition scheme. Iwashita et al. [10] propose a method where they have a human body image divided in areas and features for each area are extracted. Sinha et al. [11] use skeleton data using Kinect, where they use Adaptive Neural Network (ANN) for feature selection and classification, while Kumar and Babu [12] proposed an algorithm which also uses 3D skeleton information and trajectory covariance of joint points. Gabel et al. [13] also present a system based on Kinect sensor. Also there are many other works which use a skeleton information.

2 Proposed solution

2.1 Sensors used, preprocessing performed

In this work is used bag of features or bag of visual words method for image classification. This task is realized using Matlab, high level technical computing language also with Computer Vision System Toolbox and Image Acquisition Toolbox for Matlab. As sensor is used Microsoft Kinect for Xbox 360 with Microsoft Windows SDK v1.8 installed on Windows 10 Operating system.

2.2 Features used

According to [14] image classification analyzes the numerical properties of various image features and organizes data into categories. Algorithms for classification usually use two phases of processing that include training and testing. At training phase characteristic properties of typical image features are isolated and unique description of

each classification category is created while in testing phase these feature-space partitions are used for classification of image features [14].

Bag of words model like said in [15] can be applied to image classification, treating image descriptors as words, while bag of visual words represent sparse vector of occurrence counts of a vocabulary of local image features and can be described as histogram of visual words. Figure 1 shows steps of bag of features.

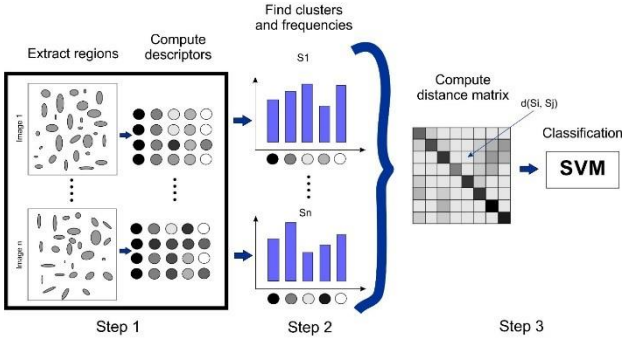

Fig. 1. Steps of Bag of features (Illustration according [16])

In Matlab can be used Computer Vision System Toolbox functions for image category classification by creating a bag of visual words. In this process it is generated a histogram of visual word occurrences that represent an image and they are used to train an image category classifier [17].

Object bagOfFeatures defines a features (visual words) using k-means clustering algorithm on extracted feature descriptors. There are iteratively grouped descriptors in k mutually exclusive clusters, where each cluster center represents a feature or visual word [17].

In this work bag of features is created as:

```
bag = bagOfFeatures(imset,'VocabularySize',500,...
   'PointSelection','Detector');
```

VocabularySize is defined by default value (500) and corresponds to K in K-means clustering algorithm, while PointSelection is selection method for picking point location for SURF (speeded up robust feature) feature extraction. Two stages exist for feature extraction in Matlab, first is method for picking point locations (SURF-Detector or Grid) and second is extracts of features (SURF extractor for both selections methods) [17]. In this case is used Detector instead of Grid.

When is used Detector for PointSelection that means the feature points are selected using SURF detector, while in other case there is predefined grid with spacing where points are picked [17].

The imset is defined as:

```
imset = imageSet('DataSet','recursive');
```

This imset load a data from dataset.

More about above mentioned terms and generally about computer vision, machine learning and other can be found in Official pages of Mathworks in various sections, supporting documents etc.

2.3 Training and testing phase

Dataset is created and tested using Kinect sensor for a 10 persons while in this work is shown for two. This process, process of creating dataset, is automated where is only necessary insert number of persons for which will be images captured and stored in dataset. Default value for insert persons' image in dataset is two and that is a reason for using two persons' example in this work. For every person it is captured 25 images for training but can be created and other number. Restriction for capturing images for dataset is range of Kinect sensor (range of Kinect is described below). This can be overcome by reduce delay between images capturing. This (reducing delay) in some manner overcome this problem, but captured images are too similar. To completely overcome this problem, it is necessary to use some more precise and large range sensor in order to capture images with more different details. Here is every person stored in different subfolder with his name. Also it is possible to automatically naming persons in format "PersonN" (where N is 0,1,2,3,4,5…). This is illustrated in figure 2 and figure 3.

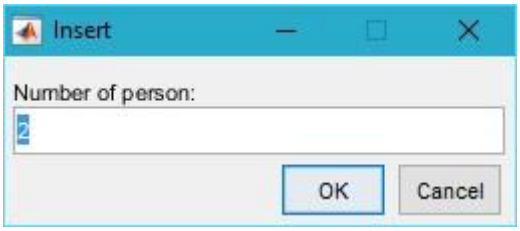

Fig. 2. Number of person for dataset

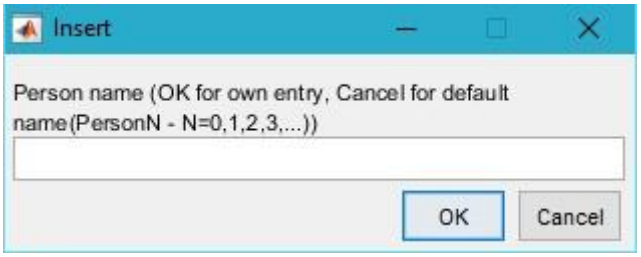

Fig. 3. Naming persons

After these operations defined in figure 2 and 3 is necessary to stand up in front of Kinect sensor and first let to Kinect detect skeleton joints of person. It is important to draw and tracking skeleton over person because skeleton is different for every person and in this case, when persons are gait recognized have important role. Joints detection and skeleton tracking over person is shown on figure 4, 5 and 6. Kinect sensor detect

and track 20 joints in standing position and 10 in sitting position. In this case is important to track all 20 joints.

Kinect sensor according [18] can recognize six person and track two.

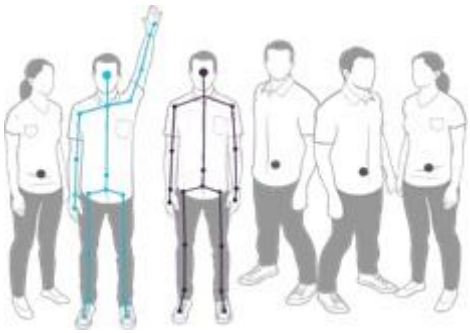

Fig. 4. Recognition of six persons and tracking two with skeleton [18]

Fig. 5. Joints detection with 3D depth sensor

Fig. 6. Skeleton tracking on a person (Also gained with Kinect)

In this work it is defined to detect and track only one person in time and also recognize one person in time. Like future work can be interesting to track and recognize more

persons simultaneously. According to [11] in default range mode Kinect can see persons standing between 0.8 meters and 4.0 meters away but practical range is between 1.2 meters and 3.5 meters while in near mode it is 0.4 meters and 3.0 meters and 0.8 meters and 2.5 meters for practical range. Guided with this, capturing images for dataset is also in that range. Images for dataset are stored in PNG (Portable Network Graphics) format along with skeleton over person. Figure 7 shows capturing images for dataset. In this case it is example for two persons like is said before. For purpose of capturing images for dataset and skeleton tracking are used both RGB Camera and 3D Depth sensor of Kinect.

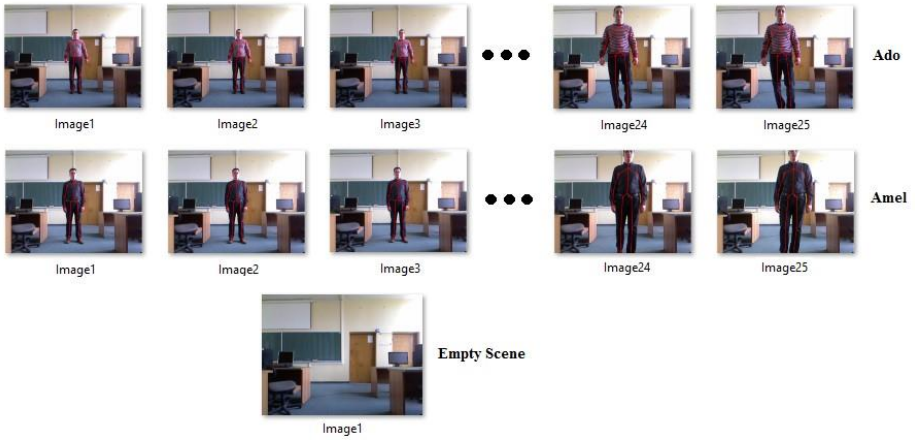

Fig. 7. Capturing images for dataset

Whole process can be described as fallows. After is dataset created using Kinect sensor than is necessary to load this dataset. This is done using above mentioned imset. Then is a step of creating bag of features from dataset of images. After that, images are encoded as a new features and then are used as training data in Matlab Classification Learner to train a model. In Classification Learner is used SVM Classifier (Support Vector Machine). Also is used and KNN (k-Nearest Neighbors) Classifier for testing purpose. Results in this case for both SVM and KNN are the same and it is 100% (Figure 10). As the both have 100%, SVM is chosen for the reason of slightly better results of using in similar cases (by our experiences).

This all procedure is a train procedure. Beside train procedure it is a prediction procedure which is reflected in real time persons' recognition. In this procedure is used trained model to give a prediction.

Figure 8 shows a steps for a presented method for people identification.

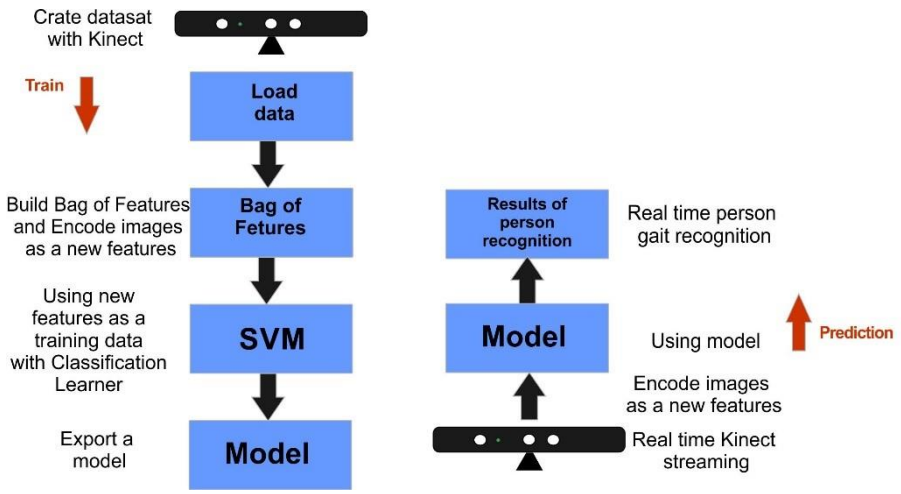

Fig. 8. Steps for a people identification

2.4 Experimental results

Results of recognition are for two persons for which is dataset captured. While is real time recognition first is shown an empty scene (Scene without persons). If a person stands in front of Kinect sensor and start walking like on figure 9, two pictures below (pictures with skeleton) on figure 9 are shown and label with name appears (on pictures above). In case if person walk away from Kinect range, again Empty Scene label appears and pictures with skeleton disappears. Table 1 shows Confusion Matrix.

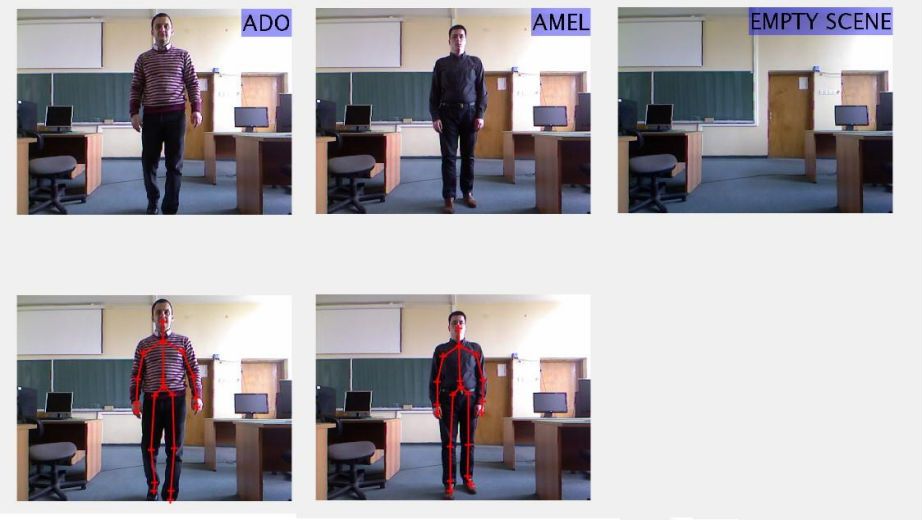

Fig. 9. Results of person recognition

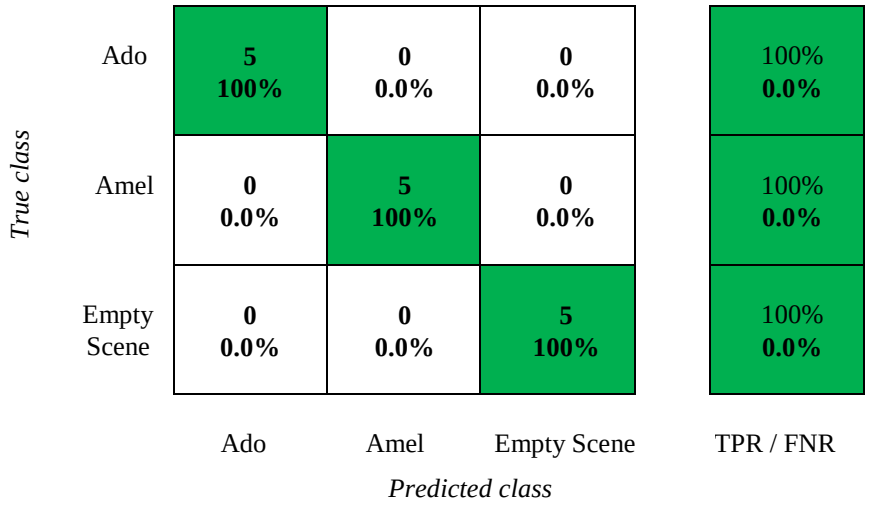

Table 1. Confusion Matrix

Table 1 shows Confusion matrix in which can be seen that prediction in all cases is 100%. This is small set of data to provide some good conclusions and much more persons must be included in experimental tests to gain real performances of the algorithm.

The algorithm is tested on small set of data, what is a shortcoming, but with these data it works very well. Also, this algorithm has some other shortcomings which are:

- The impossibility of simultaneous recognition of several persons,
- Imperfectly identification of persons from reason that images with similar characteristics can lead to wrong identification,
- The algorithm uses whole pictures with all the content, for identification (environment, people, clothes) therefore ideally works when using the same environment and the same clothes with a person which passing a process of identification.

3 Conclusion and future work

In this work, we are presented one method for people identification which uses a Kinect sensor. Many methods today are used for a process of people identification. Idea in this work is to implement an algorithm for people identification in gait because this kind of identification does not require interaction with a person. Presented algorithm works very well in some environment, but does not uses strictly analyzing of gait features of a person. The algorithm works analyzing whole pictures and features from dataset and then in real time performs prediction based on trained data. As a future work is planned to investigate and find some ways for analyzing gait features of a person and create an algorithm which will be more robust to an environment and clothes changes. Also, it is important to provide a manner for a simultaneously identification many persons.

References

1. S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “Gait energy volumes and frontal gait recognition using depth images,” in International Joint Conference on Biometrics (IJCB 2011), Washington DC, 2011, pp. 1-6.
2. S. Sivapalan, D. Chen, S. Denman, S. Sridharan, and C. Fookes, “The Backfilled GEI - A Cross-Capture Modality Gait Feature for Frontal and Side-View Gait Recognition,” in International Conference Digital Image Computing Techniques and Applications (DICTA 2012), Fremantle, WA, 2012, pp. 1-8.
3. M. Hofmann, S. Bachmann, and G. Rigoll, “2.5D gait biometrics using the Depth Gradient Histogram Energy Image,” in IEEE 5fth International Conference Biometrics: Theory, Applications and Systems (BTAS 2012), Arlington, VA, 2012, pp. 399-403.
4. R. Borràs, À. Lapedriza, and L. Igual, “Depth information in human gait analysis: An experimental study on gender recognition”, in 9th International Conference Image Analysis and Recognition (ICIAR 2012), Aveiro, Portugal, 2012, pp. 98-105.
5. J. Lu, G. Wang, and P. Moulin, “Human Identity and Gender Recognition from Gait Sequences with Arbitrary Walking Directions”, in IEEE Transactions on Information Forensics and Security, vol. 9, no. 1, 2014, pp. 51-61.
6. P. Chattopadhyay, A. Roy, S. Sural, and J. Mukhopadhyay, “Pose Depth Volume extraction from RGB-D streams for frontal gait recognition”, in Journal of Visual Communication and Image Representation vol. 25, no. 1, 2014, pp. 53-63.
7. P. Arora and S. Srivastava, “Gait recognition using gait Gaussian image”, in 2nd International Conference Signal Processing and Integrated Networks (SPIN 2015), Noida, India, 2015, pp. 791-794.
8. J. Preis, M. Kessel, and M. Werner, “Gait recognition with Kinect”, in 1st international workshop on Kinect in pervasive computing, New Castle, UK, 2012, pp.1-4.
9. A. M. Nambiar, P. Correia, and L. D. Soares, “Frontal gait recognition combining 2D and 3D data”, in Proceedings of the on Multimedia and security (MMSEC 2012), Coventry, UK, 2012, pp. 145-150.
10. Y. Iwashita, K. Uchino, and R. Kurazume, “Gait-Based Person Identification Robust to Changes in Appearance”, in Sensors, vol. 13, no. 6, 2013, pp. 7884–7901.
11. A. Sinha, K. Chakravarty, and B. Bhowmick, “Person identification using skeleton information from Kinect”, in Proceedings of the 6th International Conference on Advances in Computer-Human Interactions (ACHI 2013), Nice, France, 2013, pp. 101-108.
12. M. S. Naresh Kumar and R. Venkatesh Babu, “Human gait recognition using depth camera: a covariance based approach”, in Proceedings of the 8th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2012), Mumbai, India, 2012.
13. M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, “Full body gait analysis with Kinect”, in 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2012), San Diego, CA, 2012, pp. 1964-1967.
14. The University of Edinburgh, School of informatics, http://homepages.inf.ed.ac.uk/rbf/HIPR2/classify.htm

15. Radu Tudor Ionescu, Marius Popescu, Cristian Grozea, „Local Learning to Improve Bag of Visual Words Model for Facial Expression Recognition “, http://deeplearning.net/wp-content/uploads/2013/03/VV-NN-LL-WREPL.pdf
16. Center for Research in Computer Vision, University of Central Florida, http://crcv.ucf.edu/courses/CAP5415/Fall2012/Lecture-17-BagOfWords.pdf
17. Official page of Mathworks, http://www.mathworks.com/help/vision/ug/image-classification-with-bag-of-visual-words.html, http://www.mathworks.com/help/vision/ref/bagoffeatures-class.html
18. Official page of MSDN Microsoft, https://msdn.microsoft.com/en-us/library/hh973074.aspx