Research on human activity recognition based on image classification methods

Aistė Štulienė
Faculty of Informatics
Kaunas University of Technology
Kaunas, Lithuania
e-mail: aiste.stuliene@ktu.edu

Agnė Paulauskaitė-Tarasevičienė
Department of Applied Informatics
Faculty of Informatics
Kaunas University of Technology
Kaunas, Lithuania
e-mail: agne.paulauskaite-taraseviciene@ktu.lt

***Abstract*–Human activity recognition is a significant component of many innovative and human-behavior based systems. The ability to recognize various human activities enables the developing of intelligent control system. Usually the task of human activity recognition is mapped to the classification task of images representing person's actions. This paper addresses the problem of human activities' classification using various machine learning methods such as Convolutional Neural Networks, Bag of Features model, Support Vector Machine and K-Nearest Neighbors. This paper provides the comparison study on these methods applied for human activity recognition task using the set of images representing five different categories of daily life activities. The usage of wearable sensors that could improve classification results of human activity recognition is beyond the scope of this research.**

Keywords–activity recognition; machine learning; CNN; BoF; KNN; SVM

I. INTRODUCTION

Recently the human activity recognition problem has become a significant matter of research. In most of the cases it has a very explicit practical applicability: human activity recognition is an integrate part of human behavior-based system. Nowadays, smart home technologies are getting a lot of attention because of better care of the residents which is extremely important for elderly, children or disabled people [1]. Smart home solutions, health monitoring equipment, surveillance systems can be indicated as the typical examples of such kind of systems [2], [3], [4]. Nevertheless, there is a huge variety of specific application areas, namely anomalous behaviour detection, unhealthy habits prevention or condition tracking [5].

Nowadays, the primitive human activity partition to the static postures and dynamic motions is not sufficient. One of the key features of smart system technologies' task for human activity recognition is enabling to identify the current activity considering to the wide range of provided indoor activities. Fully-autonomous and barely noticeable assisting systems are becoming more appropriate for daily use than equipment based on wearable sensors or appliances [6], [3]. Accelerometers, gyroscopes and magnetometers have been substantiated as the most informative sensors in the sensor based recognition systems [7], [8]. Such techniques as radar, I/R or microwave, depth cameras have been widely used to obtain images [9], [10]. The commercial products such as the Nintendo's WII or Microsoft's Kinect are good examples of such devices [11]. Although these products have been partially successful, their deployment is not practical, limiting the mobility area of the human (e.g., public areas are excluded). Furthermore the wearable motion sensors make human's movement cumbersome. Additionally, the installation and maintenance of the sensors usually cause high costs. According to these facts, the more practical solutions rely on the combination of video monitoring devices and image classification methods.

Various machine learning technologies are applied for image recognition tasks. Therefore, the major challenge in human activity recognition is to evaluate the reliability of selected technologies. Considering this fact, it is necessary to compare the experimental results obtained using different machine learning approaches. In this paper, four different methods have been chosen for experiments: Convolutional Neural Networks (CNNs), Bag of Features (BoF), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). Using the same set of images representing human daily life activities these methods have been applied for the image classification into five categories.

II. IMAGE CLASSIFICATION

The general schema of human activity classification using all four methods mentioned above is presented in Fig. 1.

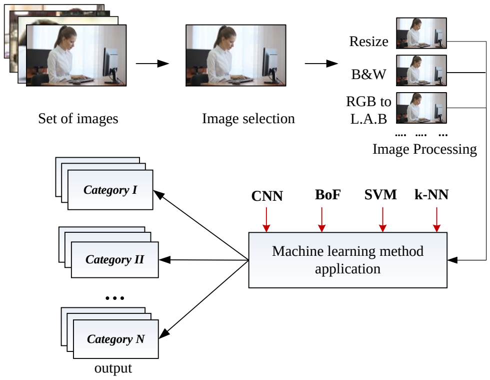

Fig. 1. The general architecture of image classification using machine learning methods

Depending on the machine learning methods, the different requirements are imposed on images. For example, using CNN, all images must be of the same size, which is usually pretty small (e.g., 224×224×3). KNN classifier may be enhanced by converting images from RGB color model to LAB model, which enables to quantify visual differences of colors and may lead to better results. SVM algorithm is used for image classification if RGB images are converted to grayscale images and then to binary images.

A. Convolutional Neural Networks

CNN is a deep learning model that obtains complicated hierarchical features via convolutional operation alternating with sub-sampling operation on the raw input images. Convolutional neural networks have become one of the most widely spread models of deep learning and have shown a very high accuracy results in various image recognition tasks [12], [13]. CNN for human activity recognition tasks usually is tested on a very popular research categories of activities (walking, jogging, running, boxing, waving and clapping) and can achieve more than 90% accuracy [14], [15]. However, in most of the cases the solutions based on CNN employ additional sophisticated sensors [16], [17]. Signals received from the accelerometer and gyroscope are transferred into a new activity image which contains hidden relations between any pair of signals. Using CNN discriminative additional features suited for human activity recognition are automatically extracted and learned [18].

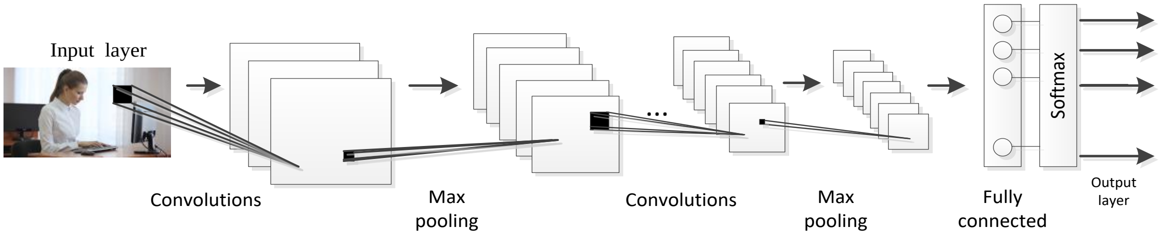

Fig. 2. A typical architecture of CNN

A general CNN architecture consists of several convolutions, pooling, and fully connected layers (Fig. 2). Convolutional layer computes the output of neurons that are connected to local regions in the input. Pooling layer reduces the spatial size of the representation to reduce the amount of parameters and computation in the network. All these layers are followed by fully connected layers leading into Softmax, which is a final classifier.

The images of the same size *a*×*a*×*b* (where *a* is the height and width of the image, *b* is the number of channels) are passed as the input to a convolutional layer. When RGB image is used, *b* is equal to 3. The convolutional layer has *m* kernels (or filters) of size *c*×*c*×*d,* where *c* is smaller than *a*.

The neurons of the convolutional layer are connected to the sub-regions of the input image (for the first convolutional layer) or the output of the previous layer. Feature map is formed when a filter moves along the input and uses the same set of weights and bias for the convolution. If *l* is a convolutional layer, the *i*th feature map $Y_i^{(l)}$ is defined using formula:

$$Y_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{m_1^{(l-1)}} K_{i,j}^{(l)} * Y_j^{(l-1)} \quad (1)$$

where $B_i^{(l)}$ is a bias matrix, $K_{i,j}^{(l)}$ is the filter connecting the *j*th feature map in layer *(l-1)* with *i*th feature map in layer *l* and $m_1^{(l-1)}$ is the amount of feature maps in layer *l-1*.

The convolutional layer is followed by an activation function. Rectified linear unit is represented by ReLU layer. ReLU is a function defined as:

$$Y_i^{(l)} = \max(0, Y_i^{(l-1)}) \quad (2)$$

$$Y_i^{(l)} = Y_i^{(l-1)}, when\ Y_i^{(l-1)} \geq 0 \quad (3)$$

$$Y_i^{(l)} = 0, when\ Y_i^{(l-1)} < 0 \quad (4)$$

Cross channel normalization (local response normalization) layer follows ReLU layer. This layer replaces all elements with normalized values. The normalized value x' for each element *x* is defined as:

$$x' = \frac{x}{(K + \frac{\alpha * s}{windowChannelSize})^{\beta}} \quad (5)$$

where *K*, *α* and *β* are hyper-parameters in the normalization, *s* is the sum of squares of the elements in the normalization window [19]. The expression can be detalized:

$$b_{x,y}^{(i)} = \frac{a_{x,y}^{(i)}}{(K + \alpha \sum_{j=\max(0,i-\frac{n}{2})}^{\min(N-1,i+\frac{n}{2})} (a_{x,y}^{(j)})^2)^{\beta}} \quad (6)$$

where $b_{x,y}^{(i)}$ is the response-normalized activity, $a_{x,y}^{(i)}$ is the activity of a neuron computed by applying kernel *i* at position *(x,y)* and then applying the ReLU nonlinearity, *n* represents adjacent kernel maps at the same spatial position, *N* is the total number of kernels in the layer.

Pooling layers follow convolutional layers and summarize the outputs of near groups of neurons in the same kernel map. The neighborhoods summarized by adjacent pooling units do not overlap. Max-pooling layer returns the maximum values of the input‘s rectangular regions and respectively, average-pooling layer returns average values.

The convolutional layer is followed by a particular amount of fully connected layers. The aim of the convolutional layer is to determine large patterns using the combinations of the features known from previous layers. In order to classify the images, the last fully connected layer combines the identified patterns. The final fully connected layer is followed by Softmax layer and classification (output) layer. In the classification layer, the network takes the values from the Softmax function and assigns each input to one of classes.

Three of CNN architectures have been selected for experiments in this paper: AlexNet [19], CaffeRef [20] and VGG [21]. These architectures have the same number of layers, but different input requirements for image size. AlexNet and CaffeRef require the size of 227×227×3, and VGG accepts the size of 224×224×3. The first convolutional layer filters the input 227×227×3 image with 96 kernels of size 11×11×3 when AlexNet or CaffeRef are used and 64 kernels of size 11×11×3 when VGG is used. The second convolutional layer uses the kernels of size $5\times5\times d$, where d is equal to 48 for AlexNet and CaffeRef and 64 for VGG architecture. Further layers filters the inputs with m kernels of size $3\times3\times d$, where d is increasing, however the exact number of d and m depends on the selected architecture.

B. Bag of Features

Bag of Features encodes the image features into a representation suitable for image classification. This technique is also often referred to as Bag of Words, because it uses image features as visual words represented as image. The features (which sometimes can be general, such as color, texture or shape) are used to find the similarities between images (Fig. 3).

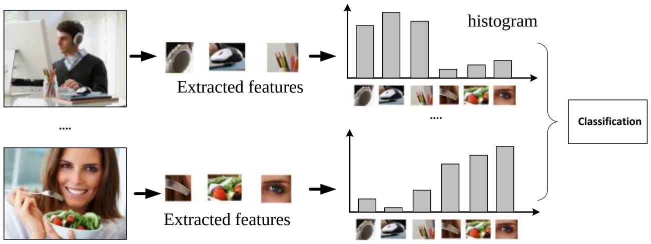

Fig. 3. Bag of Features for image recognition

BoF has shown the promising results (over 80% of accuracy) in the tasks of action recognition in video sequences [22], [23]. The typical group of sport type activities (jumping, walking, running) is used to evaluate the performance of BoF, proving that the better accuracy results can be achieved in combination with other classification methods or additional techniques [24], [25].

C. Support Vector Machine

Support Vector Machine (SVM) belongs to the class of machine learning algorithms called kernel methods. It is one of the best known methods in pattern classification and image classification. The SVM method was designed to be applied only for two-class problems. In the context of human activity classification problem, usually there are more than two possible classes (categories). Depending on this fact, it is extremely important to use modified SVM, which can be applied for multiclass classification. Two main approaches have been suggested to solve this problem [26]. The first one is called “one against all”. In this approach, a set of binary classifiers is trained to be able to separate each class from all others, where resulting class is with the highest score. The second approach is called “one against one”. In this approach the resulting class is obtained by majority vote of all classifiers.

For recognition of very simple Daily Living activities (siting, standing, walking) by carrying a waist-mounted smartphone with embedded inertial sensors, multiclass SVM (“one against all” approach) has shown an overall accuracy of more than 90%. However, the accuracy results are much lower (71.63%) trying to classify more complex activities [27]. Similarities between different actions can be explained with matched features in different sequences of actions (it may appear that running for some people is similar to the jogging for the others). However, employing 3D trajectories of body joints obtained by Kinect can provide remarkably good results of accuracy 90.57% [28].

D. K-Nearest Neighbors

K-Nearest Neighbors approach is a machine learning algorithm, which is often used for classifying objects based on the most similar training samples in the feature space. The classification is based on distance between a set of input data points and training points. Various metrics can be used to determine the distance (Euclidean distance, Mahalanobis distance, Spearman distance and etc.). KNN search enables to find k closest points in A (a set of n points) to a set of query points, when A and distance function are given. This algorithm is widely used in image processing and classification tasks.

The objects are classified according to the features of its k nearest neighbors by majority vote. Training process consists of storing feature vectors and labels of the training images. During the classification, the unlabelled query point is simply assigned to the label of its k nearest neighbors.

The performance of KNN application for classification of human activities particularly was examined using uni-axial sensors (sternum, wrist, thigh, and lower leg) [29]. Other studies based on KNN for human activity recognition have also shown rather good results of accuracy (> 90%). However it can be concluded that high accuracy results of human activity recognition based on this method can be achieved if the additional equipment (i.e., wearable sensors) is used [30].

E. Accuracy Evaluation

Human activity classification results for the particular method are often represented as confusion matrix M_{nxn} (n is equal to the amount of categories). Confusion matrix is such that the element M_{ij} is the amount of instances from category i that were actually classified as category j [6].

TABLE I. The confusion matrix for binary classification

		Predicted category	
		NO	YES
Actual category	NO	TN	FP
	YES	FN	TP

The confusion matrix for binary classification contains four elements (TABLE I): True Positives (TP) represent the amount of positive instances that were classified as positive; True Negatives (TN) represent the amount of negative instances that were classified as negative; False Positives (FP) represent the amount of negative instances that were classified as positive; False Negatives (FN) represent the amount of positive instances that were classified as negative.

The accuracy is widely used metric for the generalization of classification results. This metric is defined using formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FP} \quad (7)$$

The confusion matrix and the accuracy can also be used for *n* categories, where *n* should be more than 1 (see TABLE II – TABLE VII). In this case, the instance could be positive or negative according to the particular category, e.g., positives might be all instances of category II (e.g., sleeping) while negatives would be all instances other than category II (e.g., other than sleeping).

III. Experiments

A. The Categories of Human Activities

Despite of considerable amount of scientific research, human activity recognition only from images is still a very challenging task due to the background clutter, viewpoint, lighting, appearance and the rest of wide range aspects. Moreover, the similarities between different human actions make the classification even more challenging. The same activity may be expressed by people who have completely different appearance, body movements, postures and habits [31]. These criterions affect the way how people perform the particular action, consequently it becomes quite complicated to define the activity. Changing lifestyle, small or modern accommodations affect the employment of home areas: the rooms are usually used not only by their primary purpose (e.g., the resident can work with computer in the kitchen or eat in the bedroom). Home appliances, computers, mobile devices and other stuff around the person in most of the cases are not connected with the current activity. Even if the resident uses them at the moment, due to the changing technologies and trends they can be barely noticeable or recognizable. Considering the unsolved human activity recognition problems based on image classification methods, further theoretical and practical studies need to be carried out in order to improve the results or reject inadequate solutions.

In this paper the experimental scenario including five possible categories depending on the type of human activity has been created (Fig. 4). The activities are supposed to be performed in home or office areas. Category I relates to the situation when the people are communicating. Category II is assigned to the situation when the people are sleeping or having a rest. Category III represents empty spaces (human staying temporarily in the selected area). Human's work at computer, reading, writing or studying is assigned to category IV. Any type of eating or drinking activities are assigned to the category V. Differently from common activities' images in various recognition tasks, images representing these activities include all other objects naturally appearing while performing the particular activity. Therefore, the accuracy of expected results may not be as high as they are provided in previous researches (especially where additional techniques or methods are included).

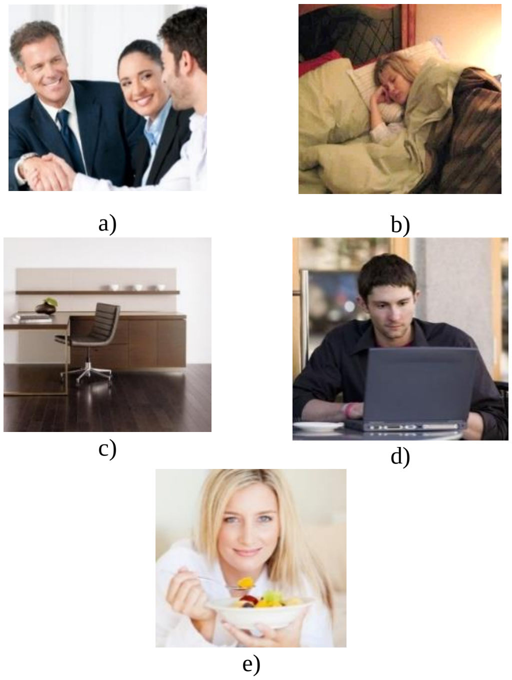

Fig. 4. Examples of five different categories of human activities: a) represents category I, b) category II, c) category III, d) category IV and e) category V.

B. Experimental Results

The image datasets containing 502 images for each category of human activity have been collected. Each dataset has been split into a training set (which contains 400 images for each category) and a test set (which contains 102 images for each category). The data for training and testing has been chosen randomly from the primary datasets.

The experiments of human activities' classification have been implemented using MATLAB software and Add-Ons [32]. The implementation of CNNs has been accomplished using MatConvNet [33], which is an open source implementation of CNNs in MATLAB environment. There also exist specific software and hardware requirements for the implementation of CNNs, such as MATLAB 2015a (or later version), C\C++ compiler, the computer with CUDA – enabled NVIDIA GPU with compute capability 2.0 or above.

The estimated classification accuracy of human activity recognition task using different image classification methods is presented in TABLE II – TABLE VII. The average accuracy of KNN is the worst one and approaches to 40.98% (although, it is more than twice the probability to choose the correct class randomly). The difference between average accuracy of SVM and Bag of Features is less than 9%. The values are 59.61% and 68.24%, respectively (the probability to choose the correct class using one of these methods is more than 0.5). The use of CNN architectures (AlexNet, CaffeRef and VGG) provides very similar results. Despite this fact, the average accuracy of AlexNet is the best one and approaches to 90.78%.

TABLE II. Confusion Matrix Of AlexNet Architecture

AlexNet average accuracy: **90.78%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	93	4	2	3	0	102
	II.	7	87	2	2	4	102
	III.	1	1	100	0	0	102
	IV.	3	3	0	93	3	102
	V.	7	4	0	1	90	102

TABLE III. CONFUSION MATRIX OF CAFFEREF ARCHITECTURE

CaffeRef average accuracy: **88.04%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	86	9	1	3	3	102
	II.	7	91	2	2	0	102
	III.	0	1	101	0	0	102
	IV.	8	6	1	82	5	102
	V.	1	4	2	6	89	102

TABLE IV. CONFUSION MATRIX OF VGG ARCHITECTURE

VGG average accuracy: **88.43%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	91	6	1	1	2	102
	II.	8	89	3	2	1	102
	III.	1	2	99	0	1	102
	IV.	4	5	1	91	1	102
	V.	5	5	0	10	81	102

TABLE V. CONFUSION MATRIX OF BOF

BoF average accuracy: **68.24%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	71	9	8	2	12	102
	II.	17	75	7	1	2	102
	III.	6	9	82	1	4	102
	IV.	7	16	19	47	13	102
	V.	8	8	4	9	73	102

TABLE VI. CONFUSION MATRIX OF SVM

SVM average accuracy: **59.61%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	59	23	6	10	4	102
	II.	19	57	10	5	11	102
	III.	6	12	67	12	5	102
	IV.	10	8	12	58	14	102
	V.	8	9	4	18	63	102

TABLE VII. CONFUSION MATRIX OF KNN

KNN average accuracy: **40.98%**		Predicted class					Total:
		I.	II.	III.	IV.	V.	
Actual class	I.	56	13	10	13	10	102
	II.	27	34	13	14	14	102
	III.	18	13	39	16	16	102
	IV.	12	11	9	44	26	102
	V.	12	15	13	26	36	102

The experimental results have shown that activities of category III determine the best results of classification for all methods except KNN (Fig. 5).

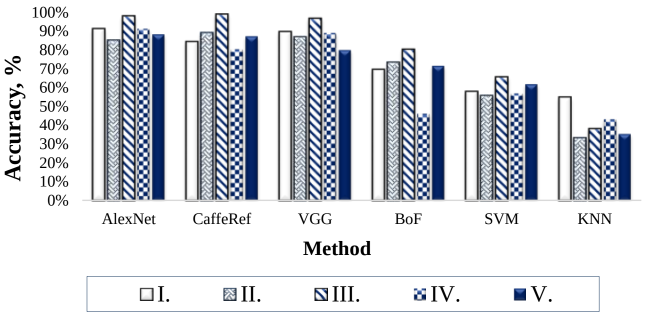

Fig. 5. Comparison of image classification methods

The recognition of activities belonging to II, IV and V categories are the most complicated, therefore provides the worst results of classification.

IV. CONCLUDING REMARKS

In this paper the research of different machine learning methods used to recognize human activities has been performed. Four different classical methods of machine learning have been selected in this research, including CNNs, BoF model, SVM and KNN. This paper provides the comparison study of the mentioned methods for human activity recognition only from images using five different categories of daily life activities. Issues related to wearable sensors or other additional techniques have not been considered. The obtained accuracy results satisfy our expectations, especially taking into account the consideration that images representing these activities include all other objects naturally appearing while performing the particular activity. The average accuracy of image classification using BoF is 68.24%. The average accuracy using SVM is lower and approaches 59.61%. Based on the experimental results we can conclude that KNN is not an appropriate method for human activity classification, using such complicated pictures of activities and applying classical KNN notation without any improvements or technological supplements. The application of different CNN architectures has revealed very similar high accuracy results, although AlexNet has reached more than 90% average accuracy, which indicates the best score of all applied methods. Considering the obtained results, further studies are needed to analyze the eligibility of different and newly created CNN architectures for the solution of image-based human activity classification problem.

REFERENCES

[1] T. van Kasteren, G. Englebienne, B. Krose, “An activity monitoring system for elderly care using generative and discriminative models,” Personal and Ubiquitous Computing, vol. 14(6), pp. 489-498, 2010.

[2] Y. Liang, X. Zhou, Z. Yu, B. Guo, “Energy-efficient motion related activity recognition on mobile devices for pervasive healthcare,” Mobile Networks and Application, vol. 19(3), pp. 303-317, 2014.

[3] J. Iglesias, J. Cano, A. M. Bernardos, J. R. Casar, “A ubiquitous activity-monitor to prevent sedentariness,” IEEE Conference on Pervasive Computing and Communications, pp. 667-680, 2011.

[4] S. Thomas, M. Bourobou, Y. Yoo, “User Activity Recognition in Smart Homes Using Pattern Clustering Applied to Temporal ANN Algorithm,” Sensors, vol. 15(5), pp. 11953-11971, 2015.

[5] X. Zhu, Z. Liu, J. Zhang, “Human Activity Clustering for Online Anomaly Detection,” Journal of Computer, vol. 6(6), pp. 1071-1079, 2001.

[6] O. D. Lara, M. A. Labrador, “A survey on human activity recognition using wearable sensors,” IEEE Communications Surveys & Tutorials, vol. 15(3), pp. 1192-1209, 2013.

[7] P. Gupta, T. Dallas, “Feature Selection and Activity Recognition System using a Single Tri-axial Accelerometer,” IEEE Trans. Biomed. Eng., pp. 1780-1786, 2014.

[8] L. Atallah, B. Lo, R. C. King, G. Z. Gitang, “Sensor positioning for activity recognition using wearable accelerometers,” IEEE Transactions on Biomedical Circuits and Systems, vol. 5(4), pp. 320-329, 2011.

[9] M. A. A. H. Khan, et al., “RAM: Radar-based activity monitor,” IEEE INFOCOM 2016, Computer Communications, pp. 1-9, 2016.

[10] A. Dubois, F. Charpillet, “Human activities recognition with RGB-Depth camera using HMM,” Conf. Proc. IEEE Eng. Med. Biol. Soc., 2013.

[11] J. Shotton, et al, “Real-time human pose recognition in parts from single depth images,” IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[12] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” Computer Vision Foundation, pp. 770-778, 2015.

[13] M. D. Zeiler, R. Fergus, “Visualizing and understanding convolutional networks,” In Proceedings ECCV, 2014.

[14] S. Ravimaran, R. Anuradha, “Survey of Action Recognition Methods for Human Activity Recognition,” In International Journal of Advanced Research in Computer Science and Software Engineering, vol. 6, pp. 284-284, 2016.

[15] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, A. Baskurt, “Sequential Deep Learning for Human Action Recognition,” In International Workshop on Human Behavior Understanding, vol. 7065, Lecture Notes in Computer Science, pp. 29-39, 2011.

[16] J. B. Yang, M. N. Nguyen, P. P. San, X. L. Li, S. Krishnaswamy, “Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition,” Proceedings of the 24th International Conference on Artificial Intelligence, pp. 3995-4001, 2015.

[17] N. Y. Hammerla, S. Halloran, T. Plotz, “Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables,” In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, pp. 1533-1540, 2016.

[18] W. Jiang, Z. Yin, “Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks,” Proceeding of the 23rd ACM international conference on Multimedia, pp. 1307-1310, 2015.

[19] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” Advances in Neural Information Processing Systems, pp. 1106-1114, 2012.

[20] Y. Jia, et al, “Caffe: Convolutional architecture for fast feature embedding”, In Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678, 2014.

[21] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Conference ICLR, 2014.

[22] M. Zhang, A. A. Sawchuk, “Motion primitive-based human activity recognition using a bag-of-features approach,” ACM symposium on International health informatics (IHI), pp. 631-640, 2012.

[23] J. C. Niebles, H. Wang, “Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words,” International Journal of Computer Vision, vol. 79(3), 2008, pp. 299-318.

[24] T. D. Campos et al., “An evaluation of bags-of-words and spatio-temporal shapes for action recognition,” IEEE Workshop on Applications of Computer Vision (WACV), 2011.

[25] M. M. Ullah, S. N. Parizi, I. Laptev, “Improving Bag-of-Features Action Recognition with Non-Local Cues,” Proceedings of the British Machine Vision Conference, pp. 1-11, 2010.

[26] C. W. Hsu, C. J. Lin, “A comparison of methods for multiclass support vector machines,” IEEE Transactions on Neural Networks, vol. 13(2), pp. 415-425, 2002.

[27] C. Schuldt, I. Laptev, B. Caputo, “Recognizing Human Actions: A Local SVM Approach∗,” Proceedings of the 17th International Conference on Pattern Recognition, 2004.

[28] M. A. Bagheri, Q. Gao, S. Escalera, “Support vector machines with time series distance kernels for action classification,” IEEE Winter Conference on Applications of Computer Vision, 2016.

[29] F. Foerster, J. Fahrenberg, “Motion pattern and posture: Correctly assessed by calibrated accelerometers,” Behavior Research Methods, Instruments, and Computers, vol. 32(3), pp. 450-457, 2000.

[30] F. Chamroukhi, S.Mohammed, D. Trabelsi, L. Oukhellou, Y. Amirat, “Joint segmentation of multivariate time series with hidden process regression for human activity recognition,” Neurocomputing, vol. 120, pp. 633-644, 2013.

[31] M. Vrigkas, C. Nikou, I. A. Kakadiaris, “A Review of Human Activity Recognition Methods,” In *journal Frontiers in Robotics and AI*, vol. 2, Article 28, 2015.

[32] R. Collobert, K. Kavukcuoglu, C. Farabet, “Torch7: A MATLAB-like environment for machine learning,” BigLearn, NIPS Workshop, 2011.

[33] A. Vedaldi, K. Lenc, “MatConvNet: Convolutional Neural Networks for MATLAB,” Proceedings of the 25th annual ACM international conference on Multimedia, pp.689-692, 2015.