AUTOMATYKA • 2011 • Tom 15 • Zeszyt 3

Zbigniew Mikrut*, Magdalena Smoleń**

A Neural Network Approach to Recognition of the Selected Human Motion Patterns

1. Introduction

Many recent efforts contributing to research and practical applications in the field of biomedical engineering concerned the home-care monitoring. The problem is of great importance, especially nowadays, when the number of elderly people in the population exhibits a permanent growing tendency and the related number of people in bad health is constantly increasing. The ADL (Activity of Daily Living) recognition is a relevant part of the evaluation of quality of life for elderly or disabled people. It enables the people requiring occasional support to feel more safe and independent, thus offering also a chance to sustain their motional activity to the old age.

In recent literature the authors propose various types of recording methods and algorithms in order to obtain significant information about the health condition and physical activity of the person under supervision.

In [6] Juang and Chang propose home care system detecting four most common body postures (standing, bending, sitting, lying) and sudden accidental falls. For the posture classification a person's silhouette is extracted from a fixed camera frame. Feature vectors are built with Fourier transform coefficients of both horizontal and vertical projection histograms and with the silhouette length-to-width ratio. The human body posture classification based on a neural fuzzy network shows high level of accordance to the experimental results and usability for home care emergency detection related to dangerous fall episodes.

Simplified methods and rules are used in [7]. The real-time system is composed of three parts: human segmentation, recognition of human posture and finally posture identification. In the second stage the authors introduce simple rules imposed on human body

* AGH University of Science and Technology, Faculty of Electrical Engineering Automatics, Computer Science and Electronics, Department of Automatics, al. A. Mickiewicza 30, 30-059 Krakow, Poland

** AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, al. A. Mickiewicza 30, 30-059 Krakow, Poland

parameters (the length and width of the selected parts of the body, for example: lower and upper body part, head, etc.). By methods used in that paper several postures can be recognized, namely: standing, sitting, kneeling and stooping.

Htike and Othman [4] study the supervised (MLP) and unsupervised (SOM, FCM and *k-means*) learning of classifiers for human posture recognition in video sequences. After the training stage, the system is evaluated by the task of classifying five chosen postures using both type of classifiers. The study shows that the supervised learning classifiers bring better results, while the unsupervised ones perform worse in cases with a high number of postures to be recognized.

In [5] authors present interesting method for vision-based human postures estimation using body silhouette and skin-color information. The problem consists in location of five significant body points (head, tips of the feet and tips of the hands) using body shape constraints. The points are chosen from convex points on a defined curve. The selection of these points is possible due to heuristic rules based on body shape characteristics.

Our paper applies to the automatic analysis of selected human motion pattern from the recorded video sequences. The motion pattern representation was obtained by means of *optical flow* histograms, while in the recognition stage artificial neural networks approach was used. Several movements were investigated in the study, such as: walking, reaching (by hand) for the object or transition between the selected body positions.

2. Recording and selection of data

In the present study the human body motion measurements were performed by digital video camera arranged at the subject's side (perpendicularly to the gait direction). The sampling frequency (recording speed) was 25 frames per second. During the studies four healthy volunteers were investigated. The task was to carry out several specified types of movements: natural gait, squatting, sitting and reaching upwards and forward for some object using one hand. Each person was asked to repeat the above activities a few times in a continuous sequence.

After data collection the next step consisted in the manual selection of the frames corresponding to a given type of activity. In consequence of this division we received an increase of the activity type sets from the initial general five to nine specific motion types (Tab. 1, Fig. 1):

1) Unbounded walking with natural speed (Fig. 1a).
2) Going from stand to squat position (Fig. 1c).
3) Going from squat to stand position.
4) Sitting down on chair (from stand to sit position) (Fig. 1d).
5) Standing up from chair (from sit to stand position).

6) Reaching upwards for some object with one hand – start:
   – start – upper limb situated perpendicular to floor, hand situated downwards (Fig. 1i),
   – end – upper limb situated perpendicular to floor, maximal range of the hand reach, hand situated upwards (Fig. 1k).
7) Reaching upwards for some object with one hand – return:
   – start – upper limb situated perpendicular to floor, maximal range of the hand reach, hand situated upwards (Fig. 1k),
   – end – upper limb situated perpendicular to floor, hand situated downwards.
8) Reaching forward for some object with one hand – start:
   – start – upper limb situated perpendicular to floor, hand situated downwards (Fig. 1e),
   – end – upper limb situated parallel to floor, maximal range of the hand reach (Fig. 1g).
9) Reaching forward for some object with one hand – return:
   – start – upper limb situated parallel to floor, maximal range of the hand reach (Fig. 1g),
   – end – upper limb situated perpendicular to floor, hand situated downwards.

Table 1

Number of the selected frames with the given type of activity (for 4 people named as K, O, C, M)

Type of activity	Person code	K	O	C	M	K+O	C+M
Walking	1	177	298	522	213	475	735
Going to the squat	2	77	150	277	626	227	903
Standing from the squat	3	67	213	362	621	280	983
Sitting down on the chair	4	210	574	319	354	784	673
Standing up from the chair	5	133	553	277	268	686	545
Reaching – upwards – start	6	161	435	265	379	596	644
Reaching – upwards – return	7	166	381	320	318	547	638
Reaching – forward – start	8	148	315	214	407	463	621
Reaching – forward – return	9	270	550	376	448	820	824
	min	67	150	214	213	227	545
	min/3	22	50	71	71	75	181

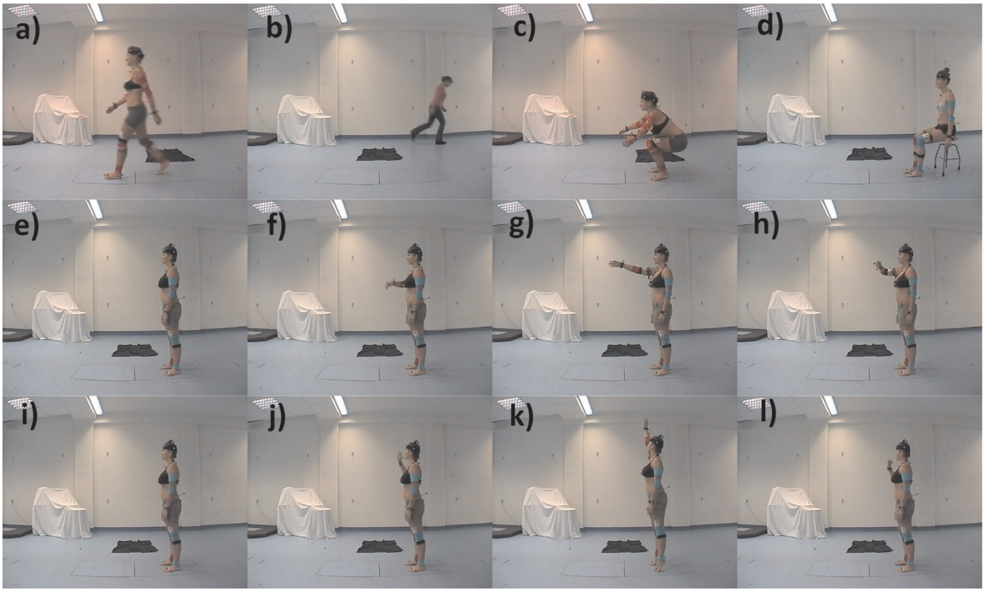

Fig. 1. Frames with examples of the selected activities: a) walking to the left; b) walking to the right; c) squatting; d) sitting; e) – h) reaching forward for an object (start and return); i)–l) reaching upwards for an object (start and return)

3. Performance of motion representation

Due to the fact that the presented research is at an early stage it was decided to use a comprehensive description of the movement of human silhouettes, based on the calculation of the optical flow. Optical flow can be calculated using different algorithms and with varying degrees of success [2, 8]: faster algorithms generate less accurate motion field, but sufficient to detect moving objects silhouettes, while the calculation time for accurate algorithms is too long to apply them in practice. The method was successfully applied in the Laboratory of Biocybernetics to detect car motion at intersections [1]. The results from the research described in [10] show that the two basic gradient methods – Horn–Schunck [3] and Lucas–Kanade [9] provide similar effects for the appropriately selected parameters.

In the presented study we applied the Horn–Schunck algorithm, which uses the first order difference as a method of numerical differentiation. The disadvantage of this differentiation method is the susceptibility to interference, the advantage – using only two consecutive frames of a sequence. The obtained computation speed was satisfactory: the optical flow of 640×480 size image is calculated within 0.1 s on the computer Intel Core i7 920, 2.66 GHz, operating under Windows7 ×64 system. To ensure the detection of the smaller movements every third frame of a movie was analyzed.

Detection of moving objects was performed by binarization of optical flow module with a constant threshold [10]. After rejecting the object part representing motion detected in the previous frame the actual silhouette was received. Figure 2a shows one movie frame with the outline of the detected silhouette drawn in red.

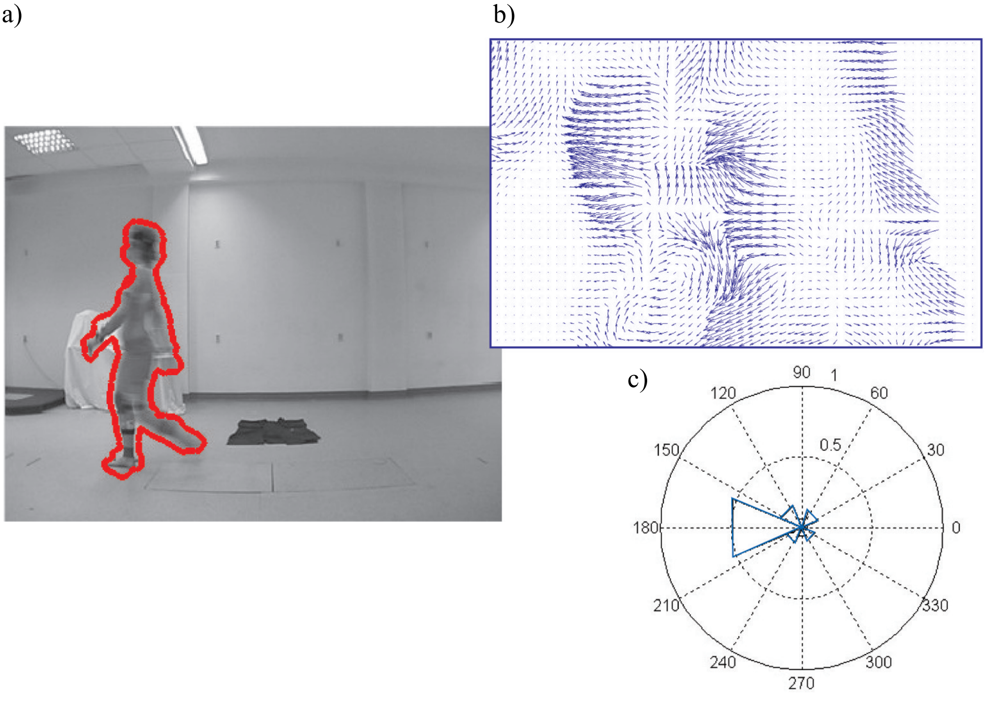

Fig. 2. Frame presenting the „natural gait" activity, with the detected object outline (a); b) optical flow in the part marked in blue in the Figure a); c) histogram of motion field directions calculated within the broaden edge of the object (polar representation)

Figure 2b shows a fragment of optical flow data, corresponding to the rectangular area marked in Figure 2a. Vectors located inside the silhouette outline have different orientations. It comes from the simplicity of the used optical flow method. However, vectors located at the silhouette edge reflect quite well the direction of object motion. Therefore it was decided that the representation of motion was formed by vectors located on the silhouette edge broadened to about four pixels. Normalized histogram of the directions of optical flow vectors aggregated to eight compartments (bins) are shown in Figure 2c. On the basis of this vector the performed activity is recognized.

Figure 3 presents representations made from the analysis of a movie consisting of 160 frames. The vertical axis shows the histogram of the aggregated optical flow directions. Eight aggregated directions (bins) correspond to the following ranges of angles [-337.5^{o} 22.5^{o}], [22.5^{o} 67.5^{o}],..., [292.5^{o} -337.5^{o}]. Angles corresponding to the centers of those ranges are respectively: 0°, 45°, ..., –315° (see Fig. 2c). On the horizontal axis the numbers of the analyzed frames divided by 3 are shown, because the movie was analyzed with 3 frame intervals. The higher value in a given histogram bin, the brighter corresponding field.

Visual analysis of Figure 3 corresponds to the real activities registered on the movie (Fig. 1a, Fig. 1b). In the first phase a person is running to the right side of the scene (range 1–13 on the horizontal axis). The direction to the right side corresponds to the first bin of the histogram (see also Fig. 2c), which are characterized by the high (white) values. In the range of 14–32 the scene is static – histogram values are equal to zero. In the range of 33–52 the investigated volunteer is marching fast to the left side. This situation is reflected on the histogram with high values of the 5th bin.

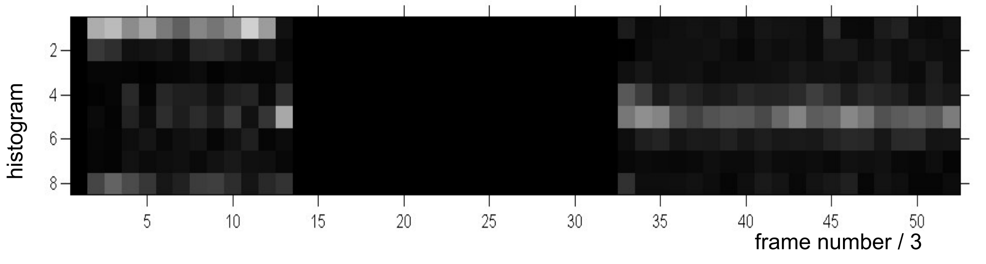

Fig. 3. Representation of the movie part by means of histograms of optical flow directions calculated for the every third frame. Motion to the right side of the scene corresponds to histogram bin equal to 1, motion to the left – bin equal to 5. Frames 1–13 – a person is running to the right, 14–32 – the scene is empty, 33–52 – a person is walking to the left

4. Classification of the activities

For identification of the motion patterns of nine different activities (see chapter 2) *backpropagation* neural network with one hidden layer was selected. It is considered as one of the best classifiers [11]. Experiments of learning and recognition stages were carried out for the structures of 8–x–9 (input layer – hidden layer – output layer). The value x was changed in the range of [3, 40].

Input data (8-elements histograms) were normalized to the [0, 1] interval and were divided into the learning and test sets respectively (approximately 1:1). Equinumerosity of the data in all training sets resulted in a reduction in the number of learning data for person K (see tab. 1). In consequence it also affected the obtained results (see Tab. 2 third row). For other people and a joint set of person C and M the amount of data for particular activities is 35 vectors. This quantity gives about 300 elements in the learning set for each person.

The summary of the recognition results for the most efficient neural networks is presented in Table 2. Recognitions of 80% were obtained for very different sizes of hidden layer. That proves the correctness of the decision to experiment with different hidden layer sizes. Testing the trained networks with the data sets from other people did not bring the expected effects (see Tab. 2 row 1). Similar recognition rates (at the level of 50–60%) were obtained for other combinations – networks trained with motion representations of one person, but tested with representations of another person.

Table 2
Best results of posture learning and recognition for each person

Nr	Person	Recognition of learning set [%]	Recognition of test set [%]	Hidden layer	Additional notes
1	C+M	96.0	82.6	33	recognition of test set K+O – 64.2%
2	C	97.8	88.1	39	---
3	K	98.9	64.8	12	too small number of learning set data
4	M	99.0	83.8	31	---
5	O	99.1	79.8	19	---

The best results were obtained for the person C. In case of five networks structures the recognition results exceeded 84% and for the 8–39–9 network reached 88% (see Tab. 3). For these networks an additional experiment was performed. It was intended to test the networks with the joint sets (learning and testing sets) from the remaining persons. The results presented in Table 3 (columns 4–6) are in the range of 50–67%, below the acceptable level of recognition, and they indicate rather high individualization level of recognized activities.

Table 3
Summary results of the selected architectures of neural networks
learned by representations of person C motion

Structure of neural network	Recognition of learning set from person C [%]	Recognition of test set from person C [%]	Recognition for person K	Recognition for person M	Recognition for person O
8-14-9	97.8	84.8	49.4	54.7	60.7
8-36-9	95.2	84.6	60.9	64.1	65.0
8-37-9	98.1	86.4	53.8	59.6	66.8
8-39-9	**97.8**	**88.1**	55.9	64.8	66.7
8-40-9	98.4	86.9	58.6	65.6	67.7

5. Summary and conclusions

In this article we propose the use of representation based on optical flow directions to recognize nine activities. Optical flow was calculated for the whole image and then the silhouette was detected by means of binarization of vectors modules. The histogram of directions of optical flow vectors was calculated inside the mask of broaden silhouette outline. Histograms created a representation of motion and were also the input for the neural networks which identified the specific activities.

The results from Table 2 prove that it is possible to identify particular activities with 80–88% rate. The test results for the networks trained with the data from other persons proved that different persons executed the activities in a different manner (Tab. 2 row 1, Tab. 3).

Table 4
Array of errors for test set of person C (network 8–39–9)

		recognized activity								
		1	2	3	4	5	6	7	8	9
expected activity	1	137	2	0	0	0	0	0	0	0
	2	0	31	0	0	2	2	0	3	1
	3	0	0	49	0	0	0	0	2	1
	4	0	0	2	46	0	0	2	0	5
	5	2	0	3	0	31	1	1	1	1
	6	0	1	0	0	1	11	0	0	0
	7	0	0	1	1	0	0	20	1	1
	8	0	0	0	0	0	0	0	6	0
	9	0	1	2	0	1	0	4	2	18

The obtained recognition results do not fully satisfy the authors. Several ways to improve these results are considered:

1. Collect more data and repeat the learning stage.
2. Take into account the possibility of rejection of the unreliable results during the recognition stage.
3. Filter in time domain the recognition results: if a person performs a particular activity, it is rather not possible that the person performs suddenly (in the next movie frame) another type of activity. As can be seen in Table 4 the errors occur separately (maximal value – 5 – is for very similar patterns: „sitting down on the chair", recognized as a "reaching forward – return").
4. Extract in the silhouettes particular parts of the body (legs, arms, head, trunk) and calculate their histograms of the motion field.

Acknowledgement

Work supported by the Polish State Committee for Scientific Research grant in years 2009–2012 as a project No. N N518 426736.

References

[1] Adamski A., Bubliński Z., Mikrut Z., Pawlik P., *The image-based automatic monitoring for safety traffic lanes intersections*. [in:] J. Piecha (Ed.), Transactions on TRANSPORT – SYSTEMS – TELEMATICS, Wyd. Politechniki Śląskiej, Gliwice 2004, 92–102.

[2] Barron J.L., Fleet D.J., Beauchemin S.S., *Performance of optical flow technique*s. Int. Journal of Computer Vision, 12(1), 1994, 43–77.

[3] Horn B.K.P., Schunck B.G., *Determining optical flow: a retrospective*. Artificial Intelligence, 59, 1993, 81–87.

[4] Htike K.K., Othman O.K., *Comparison of Supervised and Unsupervised Learning Classifiers for Human Posture Recognition*. International Conference on Computer and Communication Engineering (ICCCE 2010), Kuala Lumpur, Malaysia, 2010.

[5] Juang C.F., Chang C.M., Wu J.R., Lee D., *Computer Vision-Based Human Body Segmentation and Posture Esimation*. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, vol. 39, No. 1, 2009, 119–133.

[6] Juang C.F., Chang C.M., *Human Body Posture Classification by a Neural Fuzzy Network and Home Care System Application*. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, vol. 37, No. 6, 2007, 984–994.

[7] Li C.C., Chen Y.Y., *Human Posture Recognition by Simple Rules*. 2006 IEEE International Conference on Systems, Man, and Cybernetics, Taipei, Taiwan, 2006, 3237–3240.

[8] Liu H., Hong T., Herman M., Camus T., Chellappa R., *Accuracy vs. Efficiency Trade-offs in Optical Flow Algorithms*. Computer Vision and Image Understanding (CVIU), 72(3), 1998, 271–286.

[9] Lucas B.D., Kanade T., *An iterative image registration technique with an application to stereo vision*. Proc. 7th Intl. Joint Conf. on Artificial Intelligence (IJACAI), August 24–28, Vancouver, 1981, 674–679.

[10] Mikrut Z., Pałczynski K., *Segmentacja sekwencji obrazów z wideodetektora na podstawie przepływu optycznego*. Automatyka (półrocznik AGH), t. 7, z. 3, 2003, 371–384.

[11] Tadeusiewicz R., *New Trends in Neurocybernetics*. Computer Methods in Materials Science, vol. 10, No. 1, 2010, 1–7.