Modified kNN Algorithm for Improved Recognition Accuracy of Biometrics System Based on Gait

Marcin Derlatka

Modified kNN algorithm for improved recognition accuracy of biometrics system based on gait

Marcin Derlatka

Bialystok University of Technology,
Wiejska Street 45C, 15-351 Bialystok, Poland
m.derlatka@pb.edu.pl

Abstract. The k-nearest neighbors classifier is one of the most frequently used. It has several interesting properties, though it cannot be used in utilitarian biometric systems. This paper proposes the modification of kNN algorithm which ensures correct work even in the case of an attempt at access to the system by unauthorized people. Work of the algorithm was tested on data presenting ground reaction forces (GRF), generated during human's walk, obtained from measurements carried out on over 140 people. Differences between particular strides were determined with the Dynamic Time Warping (DTW). The recognition precision obtained was as high as 98% of biometric samples.

Keywords: k-nearest neighbors, human gait recognition, ground reaction forces.

1 Introduction

Biometrics understood as a science of recognizing a human on the basis of his/her physical and/or behavioral features becomes increasingly popular and increasingly courageously enters our life. The number of devices (e.g. computers) and institutions (e.g. banks) where a particular type of biometrics is used is still increasing. Every biometric system, irrespective of the feature this system is based on, includes attribution of a 'biometric signature' provided by a person applying for access to resources, to one of the classes registered in the data base. Obviously, if the provided biometric signature is not similar enough the class pattern stored in the data base, there should appear information on a denial of access to the resources.

Among the most intuitive classifiers are the classifiers in which similarity is defined as the inverse of the distance in the feature space, e.g. the kNN classifier. In this case the input pattern is attributed to the class which includes most of its nearest neighbors. Intuitiveness, quite good generalizing properties and easiness of implementation result in the fact that the biometric systems working the basis of the kNN algorithm are frequent in literature [9], [12]. Unfortunately, this classifier, despite its several modifications enhancing both the properties and the quality of classification [5] [12], has some features that prevent its direct application in biometric systems. They are:

- attributing a given point of the space of features to one of the classes even when the distance, though minimal, is as long as we should talk about the lack of similarity to all patterns; in this situation the desired answer of the classifier is the information of unrecognizing the given person;
- while making decisions taking into consideration the constant number k of the nearest neighbors, even if part of them are relatively far from the point under classification.

These features, occurring also among other classifiers, lead to the fact that papers connected with biometrics very often assume that we deal with a closed group of people and we are not endangered with an attempt at getting to the resources by the people from outside of this group [11]. Certainly, such an approach limits the potential application field of a given system.

One of more interesting behavioral biometrics is human gait. Gait is a very complex human activity. It is a symmetrical and repetitive phenomenon in its normal form. The movement of each limb is partitioned for the support phase - when the foot is in contact with the ground and the swing phase when the foot is lifted and moved forward. The human gait is developed during child growth. It is assumed that the human gait pattern is evaluated till the child is seven years old and next stay almost unchanged till his death. The human gait is the result of synergistic activity of: bones, muscles and nervous systems. The cooperation of those three systems makes the human movement unique for every person.

Human recognizing using gait is not a new issue. It was Cutting and Kozlowski [2] who already conducted researches which demonstrated that a human has a skill of recognizing the people he knows from a long distance by their way of moving, even if they wear clothes different than usual and have changed their hairstyle. Gafurov [4] distinguish the recent methods in human gait recognizing depending on the signals registered. They are methods based on the data obtained from:

- video;
- floor sensor;
- portable sensors.

In case of video based methods the picture registered is usually converted frame by frame into silhouette sequences. Subsequently, depending on the applied methods there occurs an attempt at reading of selected parameter's of the human's gait and classification of a person [1],[6]. Advantages of these methods are undoubtedly a possibility of free motion of the person under examination, identification of many people simultaneously and these people's approval of the presence of cameras in buildings. Problems with recognizing people in the systems based on video cameras are generated by the examined person's change of clothes, a possibility of covering the examined person by objects or other people, changes in lighting as well as the sensitivity of certain parameters sought for to the position angle of the object in relation to the camera[6]. These disadvantages do not occur in the group of methods based on the measurement of the interaction of the ground with the lower limbs of the person under examination (floor sensor based). Here the examined person must walk along an appropriately-prepared measurement path equipped with force plates [3] or a special floor with a network of photo interrupter sensors [11]. In the last method mentioned above the person under examination fully cooperates during the time of measurement. He/she is provided with measurement equipment, such as: accelerators [4], or opto-reflective markers [7].

The main objective of this paper is to demonstrate a modified kNN algorithm which can be applied in biometric systems. The work of the algorithm has been tested in author own research on human gait.

2 Ground Reaction Forces

In the biomechanical approach, the ground reaction force (GRF) is the force which is acting on the body as a response to its weight and inertia during the contact of the human plantar with the surface. The all three components of GRF are used in the presented work. They are: anterior/posterior F_x, vertical F_y and medial/lateral F_z components of GRF. The common profiles of the GRF components are presented in Fig. 1 (a-c).

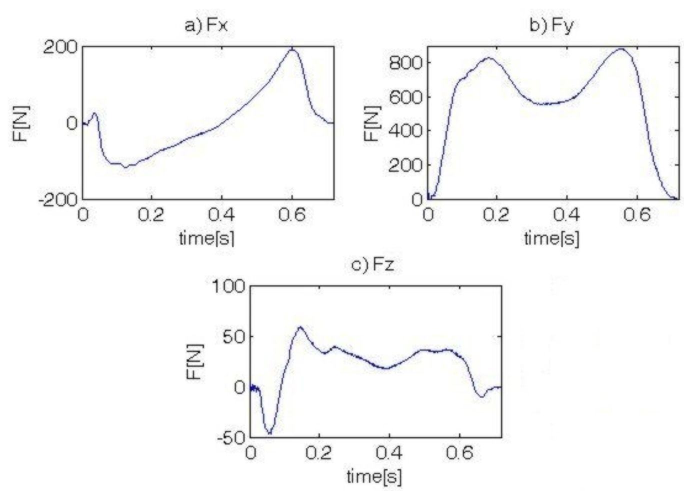

Fig. 1. Components of GRF in: a) anterior/posterior, b) vertical, c) medial/lateral direction, during the support phase of the left lower limb

The anterior/posterior component has two main phases. The value of F_x is negative in the first phase. It is a result of the deceleration of the investigated lower limb, in this case the force direction is opposite in direction of walking. The minimum of the deceleration phase is most often reached a moment before the maximum of the limb-loading phase in the vertical component of GRF. The value of F_x is positive in the second phase, respectively. The maximum of the acceleration phase is reached when the toe-off phase starts. There are three extremes in Fig. 1b. They correspond to: the maximum of the limb-loading phase, the minimum of the limb-unloading phase and the maximum of the propulsion phase (a moment before the toe off). It is not difficult to point to the same extremes for the medial/lateral component of GRF as for the vertical GRF.

3 The Modified k-NN Classification Algorithm

The k-nearest neighbors algorithm was modified for the needs of biometrics in such a way that it enables to make a decision depending on the degree of similarity between the case considered and the prototype patterns connected with a concrete user. To determine a similarity between the patterns author used a well-known DTW (Dynamic Time Warping) transform, which reduces the difference between time courses to the cost of adjusting one time series to another. The distance between n and m patterns has been calculated according to following formula:

$$D(n,m) = \sum_{c=1}^{6} D_c \tag{1}$$

where: D_c is the DTW distance between the c components of GRF for patterns n and m.

Values of similarity thresholds were defined for each user separately taking into consideration different dispersion of prototypes in the feature space for particular users. Consequently, for i-th user an average distance ρ_i between particular prototypes was calculated. The distance limit connected with a concrete threshold was calculated according to formula:

$$\rho_{\theta i} = (1.5 - \theta)\rho_i \tag{2}$$

where: θ threshold; θ=$\{0.1, 0.2, \ldots, 0.9\}$ and 1.5 is constant chosen arbitrary. It is worth noting that adopting such a definition means that for the threshold $\theta = 0.5$ all the points whose distance from the prototypes is longer than the value of average ρ_i of a given person will be treated as too dissimilar to a given class and will not be taken into account while making decisions (see point 4 of the following algorithm).

The modified k-NN classification algorithm was as follows:

1. Determine distance D of the biometric pattern under examination from all prototype patterns in the data base.
2. Select k prototype patterns whose distances D to the pattern under examination are the shortest.

3. By majority vote determine the ID of the user in the data base. If two or more users are equally numerously represented among the patterns selected in point 2 (or the remaining after rejection in point 4) - select the one whose average distance from the selected pattern is the shortest.
4. Reject k' prototypes, for which distance D is longer than $\rho_{\theta i}$ for the given threshold θ of the user whose ID was selected in point 3.
5. Check in compliance with the procedure in point 3, whether ID for K=k-k' of prototypes remained unchanged. If so, we finish the classification assigning the examined biometric pattern to the ID class. If not, return to point 3.
6. In the case when k=k' (K=0) we recognize that a given biometric pattern cannot be classified in any of the classes at the assumed threshold θ.

Such a construction of the kNN classifier enables proper classification even if near the considered i-th class point in the input space there are more prototypes appertaining to competitive (j-th) class. This is possible on condition that they are more distant from the biometric signature under consideration than the limit threshold value for j-th class and the adopted threshold θ.

4 Research Material and the Procedure of the Experiment

142 people (62 women and 80 men) took part in the research conducted in the Biaystok University of Technology. The participants in the research were 21.20 $\pm$ 1.14 years old, body weight 74.95 $\pm$ 17.0 kg and body height 174.26$\pm$ 8.81cm. During the examinations the person walked at a free speed in his/her own sport shoes on the measurement path, in which 2 Kistler's force plates were hidden, working at the frequency of 1 kHz. The volunteers performed several walks (14-20), as a result of which over 2500 strides were registered. 1056 patterns obtained from 132 people (57 man and 75 women) made a learning set (prototype points). 1500 patterns obtained from all the people under examination were treated as a testing set. The people under examination were divided into two groups: the so-called users or genuine of the system (132 people), who, as assumed, were to receive access to the system and the so-called impostors (10 people, 176 strides), who were represented only in the testing set and use to test the resistance of the biometric system to attempts at cheating (attempts at access to the data by unauthorized people).
For the obtained time courses we determined distances between all prototypes and all patterns tested in accordance with formula (1). In this paper we conducted a classification in accordance with the modified kNN algorithm proposed before, whereas for comparison we employed the classical version of the k-nearest neighbors. In both cases we assumed k=5. In the case of the classical kNN algorithm, the situation where among k-nearest neighbors was equally numerous representation of 2 or more classes, was treated as FRR error.

Table 1. Female and male characteristics

	No. of subjects	Age	Body weight	Body height
Female	62	21.15±1.10	64.47±11.68	166.83±5.53
Male	80	21.24±1.17	83.08±16.04	180.03±6.15

5 Results and Disscusion

Percentage of wrong classification for the classical kNN algorithm amounted FRR 2.19% (29 cycles) and FAR 4.00% (53 strides wrongly classified). So good an output results from a very big difference in the distances determined with DTW. The distances between the prototypes and the courses appertaining to the testing set performed by the same person were usually lower by a rank than in the case of the distances to these prototype points of strides of other people. The analysis of the situations in which an FRR error occurred in the classical kNN demonstrated that in the case of 19 strides it is possible to obtain a correct classification provided the most popular method of solving 'draws' were applied. It is enough, while making decisions, to adopt a criterion of average distance of the point under consideration to the prototypes of particular classes. Obviously, the remaining 10 strides will be classified incorrectly, which will raise the value of the FAR error up to 4.76%. It is important to note that the value of FAR error for the classical kNN is relatively low. Especially if we compare it to the results of other authors' works basing on signals of the same or similar types. Thus, for example, in the paper [9] (selected features of GRF and kNN profiles) the error amounted 7% on the group of 15 people, and in [11] (UbiFloorII and MLP): 1% and 10 people respectively. Slightly better results than in this paper were obtained in [10], where for foot pressure patterns, in the best cases, the error was 0.6% to 7.1% (104 people; 520 strides) and in [8], where for GRF and so-called wavelet packet decomposition scheme the values of the error obtained were started from 0.5% (40 people). It is important to underscore that this paper is based on much richer research material, which obviously affects the obtained results. In the case of the proposed modified kNN algorithm the value of FAR errors for the users independently from the adopted threshold are considerably lower than for the classical version of kNN. The analysis of the situations where the classical kNN generated an error allowed for demonstrating in which cases the proposed kNN algorithm works better. One of such cases occurs when a correct class has less numerous representation than a competitive class among the selected k prototypes but the distances of some of these points in the competitive class are longer than acceptable ($D>\rho_{\theta i}$). This causes rejecting a sufficient number of prototypes of the competitive class and the correct classification. In this case the classical kNN algorithm gives a wrong answer. Another possibility may occur when a correct and competitive class (classes) have an equally numerous representation among the prototypes of the shortest distance but the average distance of these points of the correct class is longer than that of the competitive one. In the situation where distances of a higher number of cases in the competitive class are longer than the acceptable kNN algorithm provides a wrong classification, whereas the proposed

Table 2. The results of the identification of the users and impostors using modified kNN

security level (thereshold)	modified kNN		
	users		impostors
	FRR	FAR	FAR
0.1	1.06%	3.25%	44.83%
0.2	1.44%	3.17%	41.38%
0.3	1.81%	2.87%	35.06%
0.4	2.27%	2.79%	29.89%
0.5	3.47%	2.57%	24.14%
0.6	4.53%	2.34%	16.67%
0.7	5.82%	1.96%	9.19%
0.8	8.23%	1.66%	7.47%
0.9	11.85%	1.06%	4.02%

one correct. Sometimes there occurs a situation where the proposed algorithm for the adopted threshold rejects all the prototypes of both the correct class and the competitive one. In this situation we obtain a FRR error. Generally speaking, it is important to state that the kNN algorithm modified for the needs of biometrics is quite conservative and such a situation occurs relatively frequently. The classifier selects the answer 'I don't know' forcing the user to repeat the trial. The value of threshold $\rho_{\theta i}$ for particular people plays a very important role. A low value of the threshold increases the value of FRR error. Along with the increase in value of the threshold the modified kNN approaches in its work a classical kNN, which reduces FRR at the cost of higher FAR. The choice of the threshold value is a compromise between the desired flexibility of the system and its resistance. Thus, taking into consideration a threshold 0.7 we force a situation where it is slightly more often than every 20 person has to repeat an attempt at access to the system. However, this value secures only every 50th trial finishing with treating user X as user Y (FAR error 1.96%: 26 wrongly classified strides). It is important to note that in this case the number of wrong classifications is more than twice lower than in the case of the classical kNN algorithm. A certain anxiety is triggered by the results of the FAR error for intruders. At least 4% (for threshold 0.7 - 9%) the acceptance of the unauthorized people is a result definitely unsatisfactory. In fact, the proposed classifier even in this case works better than a classical kNN (where the value of FAR error would be 100%); however, it indicates a need for seeking for further algorithm modifications, which will protect the biometric system from access of unauthorized people without a dramatic rise of FRR error of users.

6 Conclusion

The results obtained, basing on measurements of ground reaction forces generated during a walk conducted on an enormous, in comparison to other authors researches, data base independently from the applied classifier demonstrate a high potential of human gait as a biometric. The presented modification of the kNN algorithm for the needs of biometrics, to a considerable degree, fulfills its task. It allows for a correct classification of also those cases which have no chances with the classical version of kNN. Also a weak point of the proposed method was demonstrated: a relatively low resistance to attempts at access to resources on the part of unauthorized people. The whole encourages to continue works in this direction, which, for example more rigorous criteria of threshold $\rho_{\theta i}$ selection, should make a good method of classification still better.

References

1. Balista J. A., Soriano M. N., Saloma C. A.: Compact Time-independent Pattern Representation of Entire Human Gait Cycle for Tracking of Gait Irregularities. Pattern Recognition Letters 31, 20-27 (2010).
2. Cutting J.E., Kozlowski L.T., Recognizing Friends by Their Walk: Gait Perception Without familiary Cues. Bulletin of the Psychonomic Society, 9(5), pp. 353-356 (1977).
3. Derlatka M.: Human Gait Recognition Based on Signals from Two Force Plates. In: Rutkowski L., Korytkowski M., Schrerer R, Tadeusiewicz R., Zadeh L.A., Zurada J.M.(eds.)ICAISC'2012. LNCS, Vol. 7268: LNAI, pp.251-258,Berlin:Springer-Verlag (2012).
4. Gafurov D., Bours P., Snekkenes E.: Used Authentication Based on Foot Motion. Signal, Image and Video Processing, Vol. 5 (4), 457-467 (2011).
5. Garcia - Pedrajas N., Ortiz-Boyer D. Boosting K-nearest Classifier by Means of Input Space Projection. Expert Systems with Applications Vol. 36, 10570-10582 (2009).
6. Katiyar R., Pathak V.K., Arya K.V. A Study on Existing Gait Biometrics Approaches and Challenges. International Journal of Computer Science. Vol 10, Issue 1, No 1, 135-144 (2013).
7. Lin YC, Yang BS, Lin YT, Yang YT. Human recognition based on kinematics and kinetics of gait. Journal of Medical and Biological Engineering, Vol. 31 No. 4, 255-263 (2011).
8. Moustakidis S. P, Theocharis J.B., Giakas G.: Feature Extraction Based on a Fuzzy Complementary Criterion for Gait Recognition Using GRF Signals. In: IEEE 17th Mediterranean Conference on Control & Automation. pp. 1456-1461 (2009).
9. Orr R. J., Abowd G. D.: The Smart Floor: a Mechanism for Natural User Identification and Tracking. Proc. of Conference on Human Factors in Computing Systems (2000).
10. Pataky T.C, Mu T., Bosch K., Rosenbaum D., Goulermas J. Y.: Gait Recognition: Highly Unique Dynamic Plantar Pressure Patterns Among 104 Individuals. J. R. Soc. Interface Vol. 9, No. 69, 790-800 (2011).
11. Yun J.: User Identification Using Gait Patterns on UbiFloorII. Sensors 11, 2611-2639 (2011).
12. Zack R.S, Tappert Ch.C., Cha S-H.: Performance of a Long-Text-Input Keystroke Biometric Authentication System Using an Improved K-Nearest-Neighbor Classification Method. In: Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1-6 (2010).