# Human Identification Based on Motoric Features

Fabian Gil*, Stanisław Konatowski
Military University of Technology, Faculty of Electronics, ul. S. Kaliskiego 2, 00-908 Warsaw, Poland

## ABSTRACT

Biometric technology based on the human gait identifies people even if a person's face is covered, hidden or invisible to cameras in a dark environment. This paper presents a method of human motoric feature identification based on image recognition. One way of image recognition is described – the Haar Cascade method in conjunction with the classifier training process. Classifiers trained on MPII Human Pose and Microsoft Common Objects in Context data were used to recognize a human figure in an image. In the identification method described, joint movement parameters and characteristic body parts were analyzed. Five people were surveyed and recorded twice. The data obtained after the analysis of the first recordings, made with a camera placed at the front, served as benchmarks in the process of comparison with data from the second recordings (from behind the identified person). Data analysis was performed using a Microsoft Excel spreadsheet.

**Keywords:** human identification, motoric features, image recognition, classifiers

## 1. INTRODUCTION

In the philosophical approach, the concept of personal identity is treated as a matter of individual features; a given entity (man) made up of constantly changing properties still remains the same entity, despite these constant changes. In the process of human identification, the necessary and sufficient conditions must be met so that a given person, and the same person at another period of time, can be considered the same person [1].

The process of confirming the identity declared by a given entity is called authentication. Its purpose is to achieve a certain level of confidence that a person is in fact the one he claims to be.

The identification process is present everywhere, from opening the door to your apartment to the identification of the deceased. In the first case, the person authorized to enter the house is authenticated using the apartment key, while the second case of body identification uses a DNA genetic profile comparison, fingerprint comparison, or the victim's dentition examination.

Authentication is also used for security systems or object security. For example, this applies to the identification of a wanted person, or access to a service room. Recognition of the person authorized to enter can be carried out by comparing their access code (e.g. PIN) or characteristic features (e.g. fingerprints, retina) with data stored in a database. For wanted people, it is common practice to identify them on CCTV recordings. In this variant, this person will be compared with photos collected in the relevant archives. However, in order to positively recognize a specific person in the picture, they must be clearly visible. If the character in the recording has a covered face, you will not be able to confirm his identity.

One of the characteristics of every man is his way of moving. In the problem of recognizing people from an image, the use of the presented solution means that wearing a mask or standing at different angles to the camera to deceive the identification system will be ineffective.

---

*fabian.gil@wat.edu.pl; fax +48 261 83 74 61

# 2. CHARACTERISTICS OF SELECTED IDENTIFICATION SYSTEMS

The problem of automatic identification of people is currently gaining more and more interest, both in services that care for public security, as well as in security systems used in the workplace and private environments. When considering the issue of identification in the field of public security, we have camera systems that allow for the automatic identification of persons. The database will contain information already collected by the police or other security services. In the case of security systems in the workplace or private environments, these will be persons authorized to access a place or device. Authentication, in this case, can use one of many identification systems.

## 2.1 User ID and password

The simplest and most common form of identification of a person that allows authorized access to a given system is a password. This type of identifier is a string of limited length consisting of characters selected from a specific character set. The identifier has a user account and strictly defined permissions. Its main feature is its uniqueness within a given IT system.

In order to verify the previously declared identity (identifier provided), the system requires the user to enter a password.

## 2.2 Fingerprint reader

In recent years, the development of biometrics has made fingerprinting the most reliable method of identifying people. Comparing the scanned finger with a matrix in a database works on the principle of "finding identical points" on both matrices, not matching completely identical matrices. "Identical points" are searched by analyzing local characteristics and points of interest in the fingerprint image, as well as their mutual arrangement and slope. Local characteristics include the so-called minutia denoting the local discontinuity of the line in the form of endings, branches, and hooks. Special points include the nucleus (core) of the fingerprint and the so-called delta, which is formed in the area where the direction of the line changes. The mutual layout of the minutiae uniquely identifies the person. To recognize two fingerprints as identical, it is enough to have several common features. It is assumed that 12 common features are sufficient for defining identity.

## 2.3 Iris scanner

Iris scanning is an automated biometric identification method that uses mathematical recognition of a person's iris (from one or both eyes). Each person has a different and very complicated iris pattern in each eye. When the user registers information about his iris in a database, it is stored as an encrypted code. The user directs the eye towards the reader, and the cooperation of the IR diodes and cameras allow the iris pattern to be captured. Then they convert it into a digital version and compare it with the encrypted code in the database.

## 2.4 Face scanner

Face recognition consists of measuring the geometry of the face based on an image captured or recorded with a digital camera and comparing it with a stored pattern. The face structure consists of the eyes, nose, mouth, jaw edge and the distance between them. Each face is made up of about 80 characteristic features [2]. Based on the obtained image, a digital description of the geometry of the examined face is created. Recognition consists of comparing the description stored in the database with that created after taking the picture of the person undergoing identification or verification.

# 3. IMAGE RECOGNITION - HAAR CASCADES

The Haar Cascade is a classifier used to detect in an image the object to which it was trained, based on certain source data. Cascade training is done by applying a positive image to a set of negative images at several different stages. Better results are obtained by using high-quality images, and increasing the number of stages in which the classifier is trained.

The object detection algorithm classifies images based on simple function values that resemble Haar's base functions. Three types of functions are used to create cascades. The value of the function of two rectangles is the difference between the sum of pixels in two rectangular areas. The areas have the same size and shape, and are horizontally or vertically contiguous (Fig. 1).
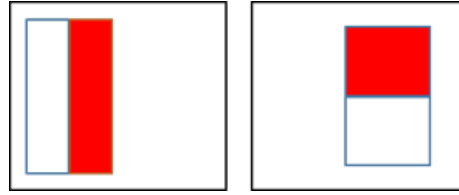
Figure 1. An example of the function of two rectangles [3].

The function of three rectangles calculates the sum in the two outer rectangles subtracted from the sum in the middle of the rectangle (Fig. 2 a). The last function - four rectangles, calculates the difference between diagonal pairs of rectangles (Fig. 2 b) [3].
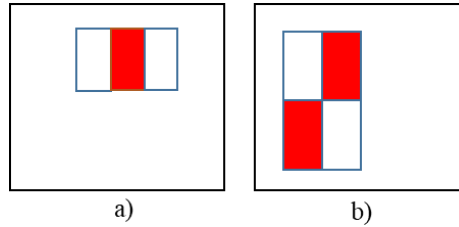


a)                                    b)

Figure 2. An example of the function of three rectangles (a) and the function of four rectangles (b) [3].

### 3.1 Classification function learning

Over 180,000 rectangular elements are associated with each image slice. This is a much larger number than the number of pixels. Even if each function can be calculated very efficiently, calculating the whole set is too expensive. The algorithm hypothesis is that a small number of these functions can be combined to create an effective classifier. In support of this hypothesis, the weak learning algorithm is designed to select a single rectangle function that best separates positive and negative examples. For each feature, a weak learning algorithm determines the optimal threshold classification function, so that the minimum number of examples is incorrectly classified. The weak classifier $h_j(x)$ therefore consists of functions $f_j$ threshold $\theta_j$ and parity $p_j$ indicating the direction of the inequality sign

$$h_j(x) = \begin{cases} 1 & for \quad p_j f_j(x) < p_j \theta_j \\ 0 & for \quad others \end{cases} \tag{1}$$

where $x$ is a sample from a 24 x 24 pixel image. In practice, no single function can perform a classification task with a low error [3].

### 3.2 Creating a cascade

By constructing a cascade of classifiers that achieves enhanced detection performance, dramatically reducing computation time, the key issue is that smaller, and thus more efficient and enhanced classifiers can be constructed that discard many negative image slices, detecting almost all positive cases. Simpler classifiers are used to discard most child windows before using classifiers of higher complexity to obtain lower false positive results [3].

The general form of the detection process is a degenerate decision tree called a "cascade" (Fig. 3). A positive result of the first classifier results in the assessment of a second classifier. A positive second classifier results in the third classifier, and so on. A negative result at any time leads to an immediate rejection of the sub-window.

The cascade structure reflects the fact that in every single image, the overwhelming majority of the windows are negative. Therefore, the cascade tries to reject as many negatives as possible at the earliest possible stage. Although a positive case will trigger a rating for each classifier in the cascade, this is an extremely rare event. Like the decision tree, subsequent classifiers are trained using these examples that go through all previous stages. As a result, the second classifier faces a more difficult task than the first. The examples that go through the first stage are "more difficult" than typical examples. For a given detection rate, deeper classifiers have correspondingly higher false positive coefficients [3].
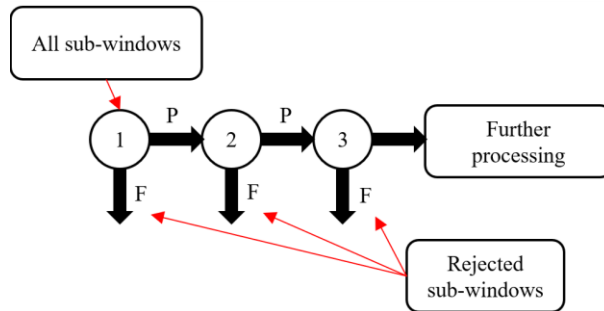
Figure 3. Schematic representation of the detection cascade [3].

## 3.3 Training the cascade of classifiers

Generally, classifiers with more functions achieve higher detection rates and lower false positive rates while classifiers with more functions require more time for calculations. In general, you can optimize the cascade and minimize the expected number of functions being processed, according to the following criteria:
• the number of stages of the classifier,
• the number of functions at each stage,
• threshold values for each stage.

Finding this optimum is an extremely difficult problem. In practice, simple structures are used to produce an effective classifier that is very efficient. Each level of the cascade reduces the false positive factor and reduces the detection rate. The target is selected for the minimum reduction of false positives and the maximum decrease in detection. Each stage is trained by adding features until target detection and false hit rates are met. Stages are added until the overall goal for the false positive and detection rates is achieved [3].

# 4. DATABASES

Until recently, progress in estimating the position of the human body has been slow due to the inability to process high-quality data sets related to pose estimation. However, new solutions represent a significant improvement in describing accurate models of the human body.

The chapter presents data sets supporting the creation of image recognition algorithms (classifiers) that detect humans.

## 4.1 MPII Human Pose

The innovative *"MPII Human Pose"* standard proposed by employees of the Max Planck Society for the Advancement of Science is an extensive data collection compiled using over 800 forms of human activity. The collection cover a wider range of human activities than previous data sets, including a variety of recreational, professional and home activities, and capturing people from a wider range of points of view. The standard provides a rich set of labels, including body point positions, full three-dimensional torso and head orientation, occlusive labels for joints and body parts, and activity labels. Adjacent video frames are provided for each image to facilitate the use of traffic information. The wealth of annotations allows you to perform a detailed analysis of the leading methods of estimating the human pose, gaining insight into the success and failure of these methods [4].

The *MPII* standard provides accurate annotations for collected images. In the photo, a person is marked and their joints, a 3D view of the head and torso, as well as the position of the eyes and nose are described. In addition, the visibility of these elements is described for all joints and marked points. The descriptions of the joints are oriented according to the position of the person in the photo, i.e. the description of the left knee will refer to his left limb [4].

## 4.2 Microsoft Common Objects in Context

The Microsoft Common Objects in COntext (MS COCO) dataset contains 91 categories of common objects, of which 82 have over 5,000. marked up labels. In total, the data set contains 2.5 million images, of which 328,000. are labeled. The

categories chosen for objects must form a representative set of all possibilities, be suitable for practical applications, and occur at a sufficiently high frequency to collect a large data set. The specificity of the category of objects can vary significantly. For example, a cat may be a member of the "mammal" or "cat" or "Persian" category. To enable the practical collection of a significant number of examples per category, it was decided to limit the data set to basic categories, i.e. category labels commonly used by people when describing objects (dog, chair, person) [5].

## 5. DESCRIPTION OF THE MEASURING STAND FOR HUMAN IDENTIFICATION

The stand for human identification consisted of two main elements. One of them was the image recording tool. The second element, responsible for processing and analyzing data, was a PC computer. Data processing took place thanks to a program written in Python, while data analysis was carried out in a Microsoft Excel spreadsheet. A simplified block diagram of the station is shown in Figure 4.

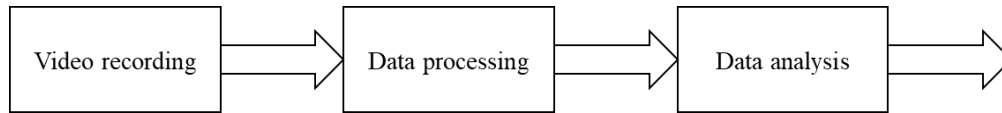Video recording → Data processing → Data analysis →

Figure 4. A simplified block diagram of the measuring stand.

### 5.1 Video recording

The recording of objects was performed in a normal perspective, i.e. with a camera placed straight in front of the man at eye level. The recorded persons moved at a constant speed along a straight line. The camera recording the movement was at a constant distance from the filmed man and was directed towards him from the front and back.

### 5.2 Data processing

The processing of data collected in the previous section was carried out by *Python* software based on the *OpenCV* graphic library and classifiers trained using databases (discussed in Chapter 4). The algorithm consisted of the following stages:
• dividing the film into individual images with a frequency of 32 frames per second,
• loading the image and measuring its parameters (width, height),
• choosing the classifier (MPII or COCO),
• adjusting image parameters to the classifier's requirements,
• passing the image through the classifier network,
• saving the coordinates of detected characteristic points,
• selecting characteristic points in the output image.

### 5.3 Data analysis

Data analysis was performed using a Microsoft Excel spreadsheet. In the first part of this process, the data resulting from image processing were properly normalized so that it was possible to juxtapose different models and compare their characteristics. Next, pattern charts were created containing selected parameters of model movement. The last step was to compare the relevant characteristics of the examined person's movement with the patterns and assess the similarity between them. The similarity was determined based on the value of the correlation coefficient (2) between the tested waveform and individual patterns, as well as the estimated dimensions of the object

$$Correlation\ coefficient(X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \tag{2}$$

where:
$X$ – set of values of the tested waveform,
$Y$ – set of reference waveform values,
$x, y$ – individual values of the sets $X$, $Y$,
$\bar{x}, \bar{y}$ – average values of sets $X$, $Y$.

# 6. RESULTS ON HUMAN IDENTIFICATION RESEARCH

Research on the system for identifying people on the basis of their motoric features was carried out under the following assumptions:

• the camera was placed at a constant height and a constant distance from a person,
• the patterns were made up of parameters from recordings of people facing the camera,
• identified people were recorded using a camera placed behind their backs,
• during identification, parameters describing head, shoulder and wrist movement, as well as human dimensions (head height and shoulder width) were used.

Five people participated in the identification system research. Each of the subjects was recorded twice during their march along the corridor at a constant speed. The first recording was made with a camera directed towards the front of the man, while the second one was made from the back. Then the recordings were divided into single frames (photos) and processed for the purpose of recognizing a person in the image, along with the detection and marking of characteristic body elements such as joints, the head, and torso. The classifiers used in the project, based on the MPII and COCO libraries, analyzed the image by estimating the location of the selected element along with the probability distribution in this area. An example of how to estimate the position of body parts is shown in Figures 5 and 6.

The area indicated by the COCO method occupies a larger area around the element being searched and has a lower probability of correct detection than the MPII method. The methods also differ in the number and type of elements searched for in the image (Fig. 7), and the time of data processing.
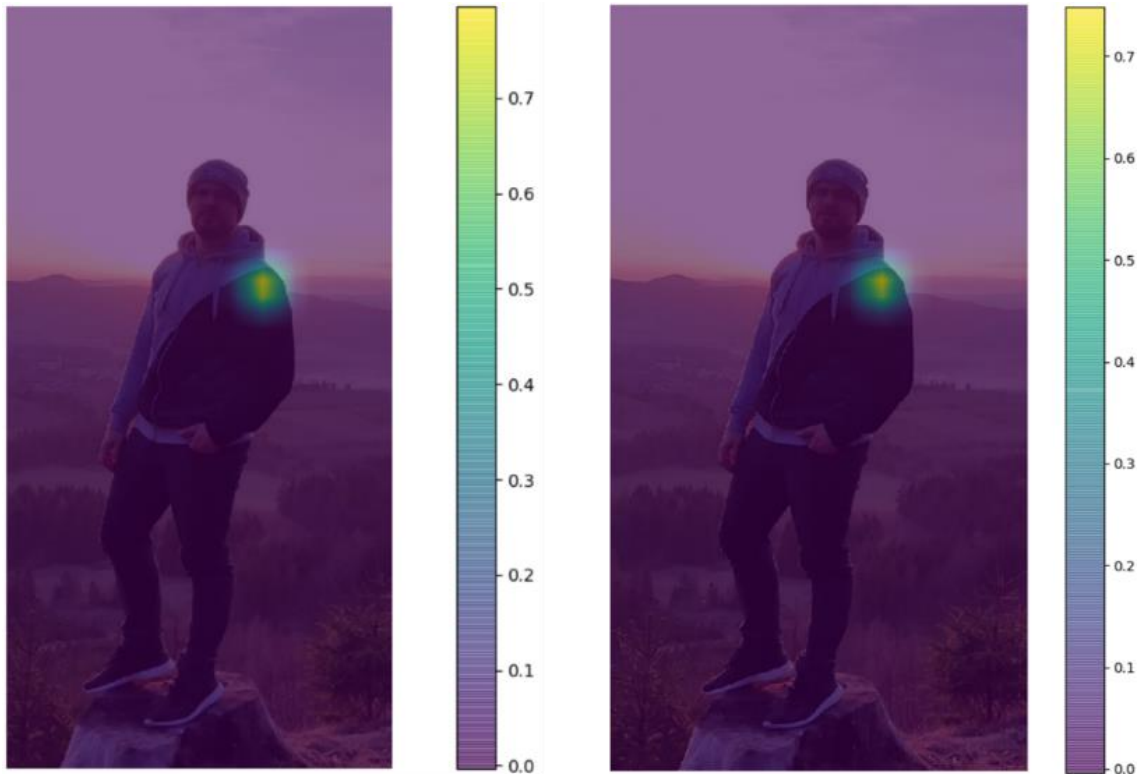


Figure 5. Estimated left shoulder position by MPII (left) and COCO (right).
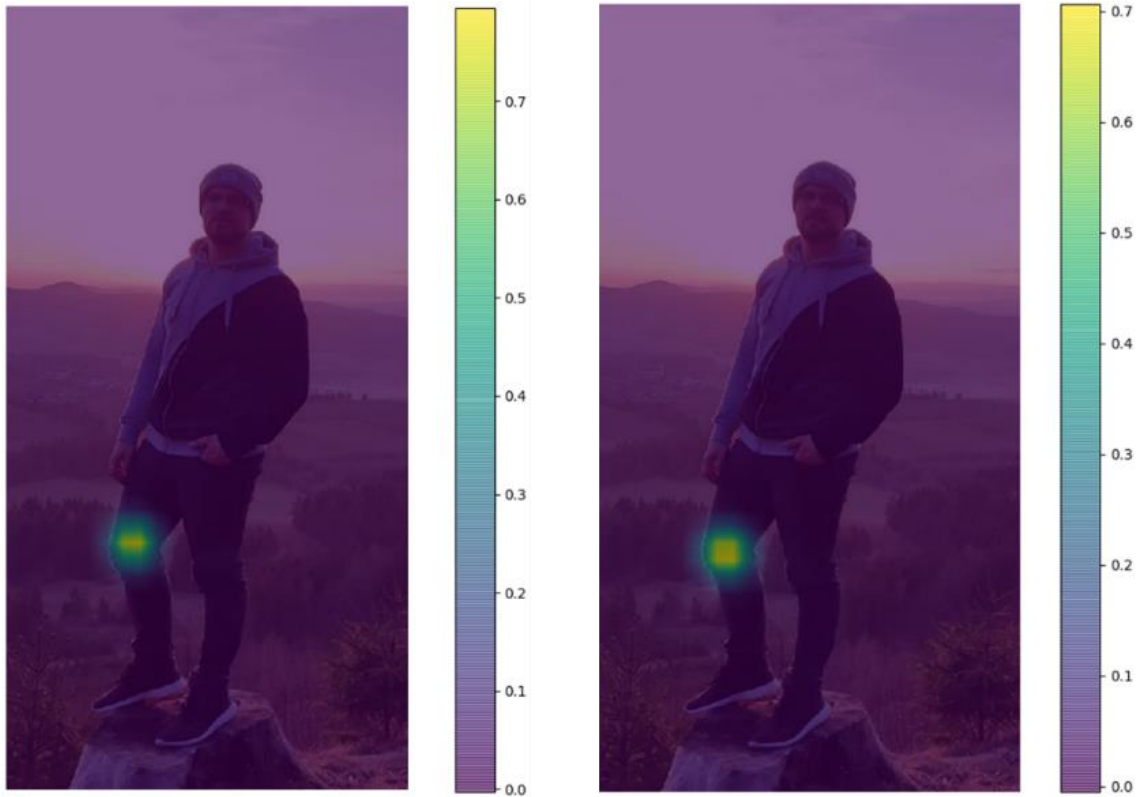
Figure 6. Estimated right knee position by MPII (left) and COCO (right).



| 0 | Head |
|---|------|
| 1 | Neck |
| 2 | Right arm |
| 3 | Right elbow |
| 4 | Right wrist |
| 5 | Left arm |
| 6 | Left elbow |
| 7 | Left wrist |
| 8 | Right hip |
| 9 | Right knee |
| 10 | Right ankle |
| 11 | Left hip |
| 12 | Left knee |
| 13 | Left ankle |
| 14 | Chest |

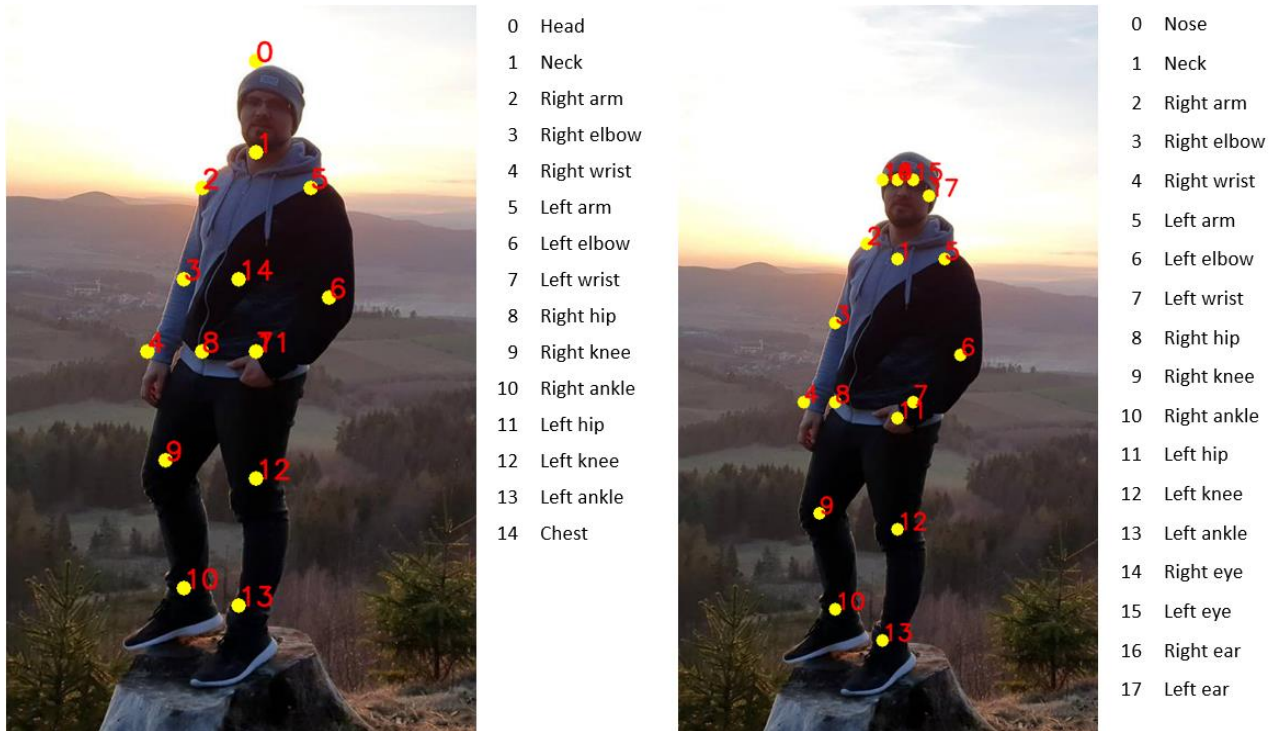| 0 | Nose |
|---|------|
| 1 | Neck |
| 2 | Right arm |
| 3 | Right elbow |
| 4 | Right wrist |
| 5 | Left arm |
| 6 | Left elbow |
| 7 | Left wrist |
| 8 | Right hip |
| 9 | Right knee |
| 10 | Right ankle |
| 11 | Left hip |
| 12 | Left knee |
| 13 | Left ankle |
| 14 | Right eye |
| 15 | Left eye |
| 16 | Right ear |
| 17 | Left ear |

Figure 7. Characteristic points according to MPII (left) and COCO (right).

Using the analyzed process for a set of images obtained from the recording, a series of data was obtained describing the coordinates of all characteristic points. Then the data were exported to a Microsoft Excel spreadsheet for further analysis.

A preliminary observation showed that both algorithms for detecting characteristic points made a lot of mistakes when interpreting the figure of a walking man. On this basis, the most stable measurements were selected for further analysis, including parameters describing the shoulders, wrists, neck and head. Models numbered 1-5 indicate people who were recorded with the camera placed in front of them, while models 6-10 characterize the same people, sorted in random order, recorded with the camera behind them. Persons 1, 2, 3, 4, 5 correspond to models 9, 8, 6, 7, 10, respectively.

First, the head movement and dimensions were examined by plotting reference characteristics describing the position of the head in the vertical plane of persons seen from the front. Data obtained with the COCO and MPII methods show similarity to each other. However, it was noticed that the first one captured the greater variability of the examined waveforms. The next step was to add to these charts waveforms describing people seen from the back. Comparing the characteristics with each other, you can probably match two of the same people seen from different perspectives. Examples of the characteristics obtained on the basis of the analysis are presented in Figures 8 and 9. A summary of the identification results by the method of assessing the correlation between the characteristics is presented in Table 1. The correctness of identification in this case was only 20%.
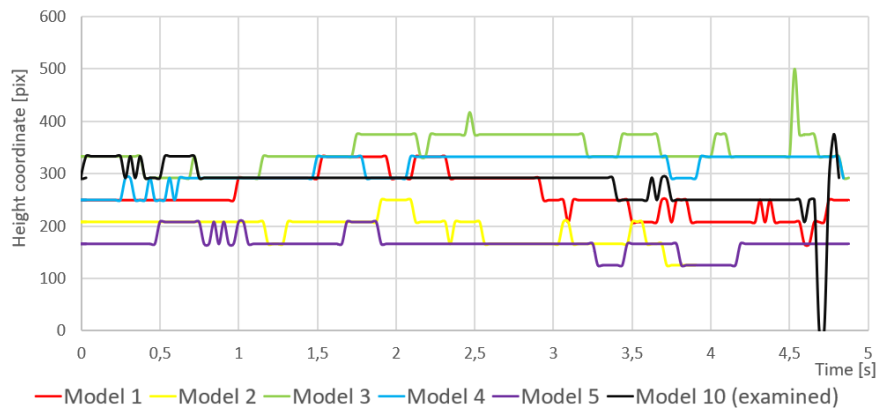


Figure 8. Comparison of the tested model with the head movement patterns - MPII method.
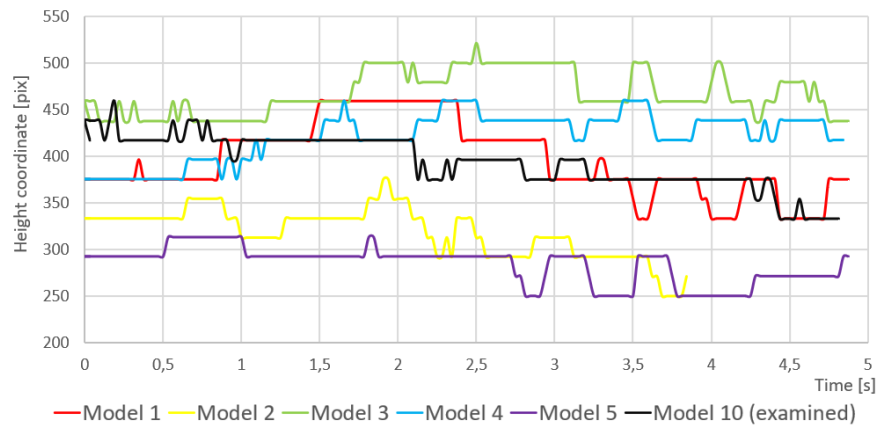


Figure 9. Comparison of the tested model with the head movement patterns - COCO method.

Table 1. Identification results based on analysis of correlation between characteristics - head movement.

| Method | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 |
|--------|---------|---------|---------|---------|----------|
| MPII | Person 2 – 51% | Person 4 – 43% | Person 1 – 56% | Person 2 – 45% | Person 2 – 43% |
| COCO | Person 1 – 30% | Person 4 – 81% | Person 4 – 72% | Person 4 – 41% | Person 2 – 70% |

Then the shoulder movement in the vertical plane was examined, considered separately for the left and right shoulder. Examples of characteristics, plotted on the basis of the analysis, are presented in Figures 10 and 11. The results of identification are presented in Table 2. These were based on the analysis of correlations between individual models and people for the characteristics presented, and the assignment of the cases to specific individuals with a percentage level of similarity and correctness of the result. As in the previous case, the recognition accuracy was classified at a fairly low level of only 40% of correct identifications. In addition, attention should be paid to the relationship between the number of incorrect identifications and the method used to recognize characteristic points on the body. Table 2 shows that much better results were obtained when working with images processed using the COCO method.
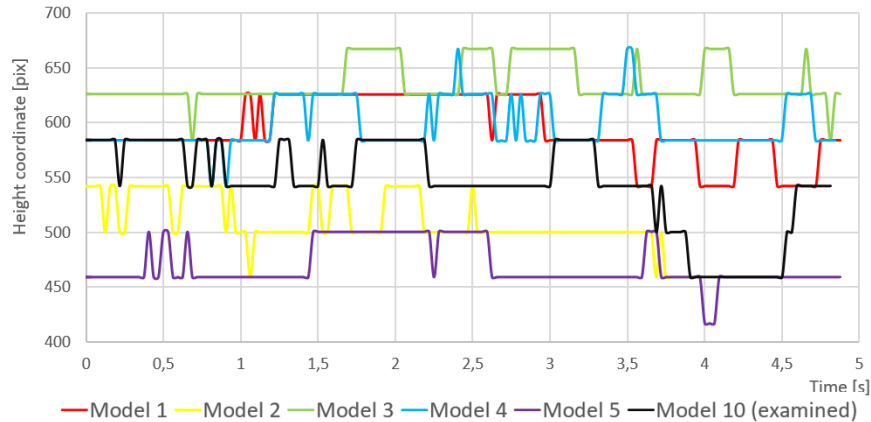


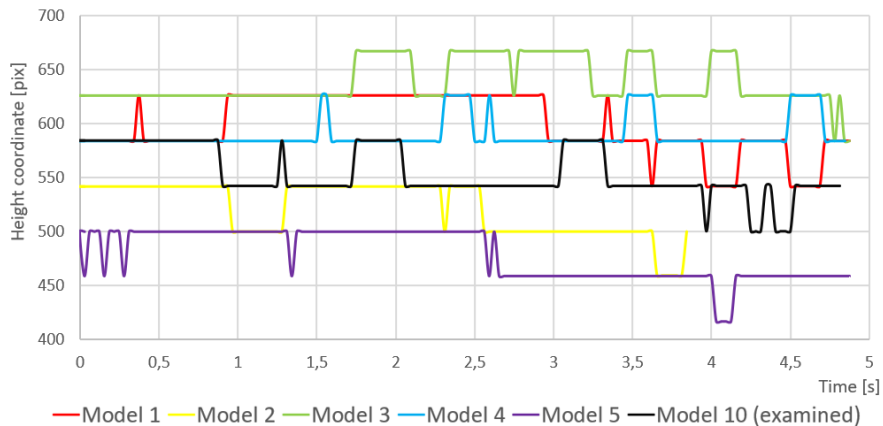Figure 10. Comparison of the tested model with the left arm movement patterns - MPII method.



Figure 11. Comparison of the tested model with the left arm movement patterns - COCO method

Table 2. Identification results based on analysis of correlation between characteristics - arm movement.

| Examined model | MPII | | COCO | |
|---|---|---|---|---|
| | **Right arm** | **Left arm** | **Right arm** | **Left arm** |
| Model 6 | Person 2 – 50% | Person 3 – 7% | Person 2 – 39% | Person 3 – 22% |
| Model 7 | Person 3 – 41% | Person 4 – 18% | Person 5 – 62% | Person 2 – 23% |
| Model 8 | Person 1 – 49% | Person 1 – 41% | Person 3 – 64% | Person 2 – 47% |
| Model 9 | Person 4 – 44% | Person 1 – 28% | Person 2 – 33% | Person 1 – 31% |
| Model 10 | Person 2 – 34% | Person 2 – 59% | Person 5 – 52% | Person 5 – 37% |

At a later stage of the examination of the position for human identification, the movement of the wrists in the horizontal plane was analyzed while moving along a straight line. Examples of the characteristics obtained on the basis of the analysis

are presented in Figures 12 and 13. Based on the analysis of the correlation between the model characteristics and the tested models, the test subjects were identified through measurements of wrist movement parameters in the horizontal plane during movement. The identification results, together with the percentage similarity level and verification of the correctness of the diagnosis, are presented in Table 3. The correctness of the identification in this case was 50%.
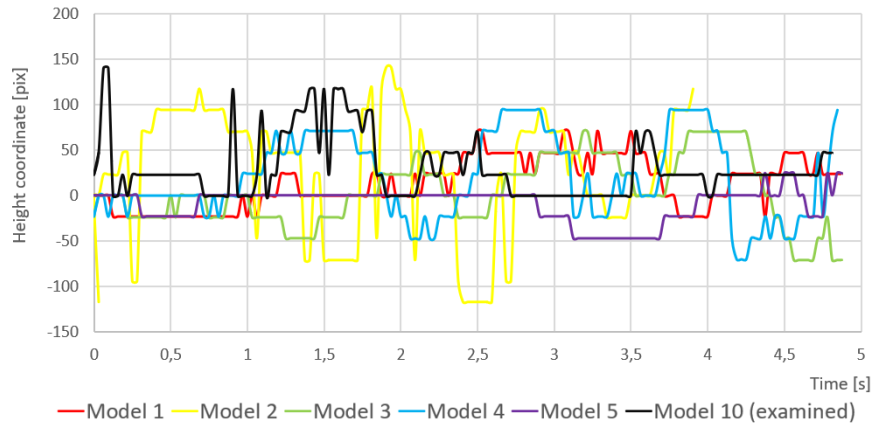


Figure 12. Comparison of the tested model with the left wrist movement patterns - MPII method.
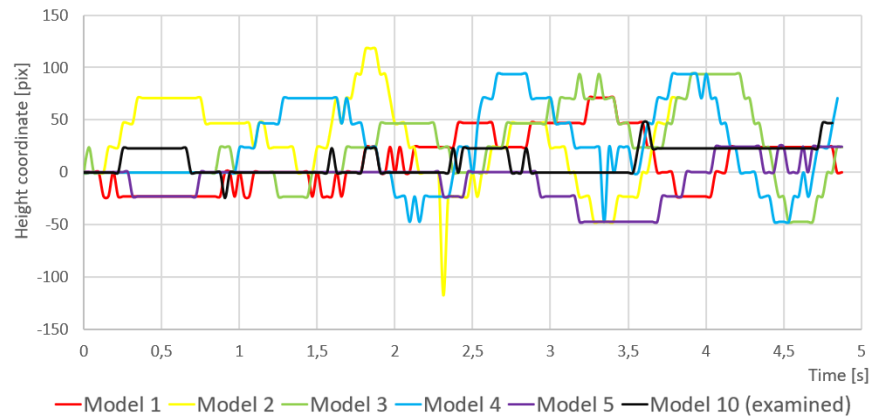


Figure 13. Comparison of the tested model with the left wrist movement patterns - COCO method

Table 3. Identification results based on analysis of correlation between characteristics – wrist movement.

| Examined model | MPII | | COCO | |
|---|---|---|---|---|
| | Right wrist | Left wrist | Right wrist | Left wrist |
| Model 6 | Person 3 – 44% | Person 3 – 32% | Person 1 – 50% | Person 5 – 28% |
| Model 7 | Person 1 – 30% | Person 4 – 57% | Person 1 – 71% | Person 4 – 36% |
| Model 8 | Person 2 – 34% | Person 1 – 26% | Person 2 – 53% | Person 5 – 37% |
| Model 9 | Person 4 – 59% | Person 1 – 42% | Person 1 – 76% | Person 1 – 71% |
| Model 10 | Person 1 – 62% | Person 3 – 39% | Person 1 – 59% | Person 5 – 16% |

# 7. CONCLUSIONS

Five people recorded twice from the front and behind were used for the study. The recordings from the front served as standards, while those recorded from behind were used in the identification process. The movement of the head, shoulders, and wrists of the subjects was analyzed. On this basis, patterns were matched to recognize people. The correctness of this action was about 40%.

The proposed system for identifying a man on the basis of his motoric traits passed the examination, despite the low efficiency of correct recognition.

The first elements of the system that could have a negative impact on the results of the research are the environment, as well as the method of creating a recording of moving people. However, on the other hand, this type of recognition was more effective than the face recognition method due to its high tolerance to image interference in the form of low lighting or low resolution. The main factors affecting the accuracy of the designed system are the methods of recognizing the human figure in the image. Already at the stage of data processing, before the analysis of movement, one could notice numerous errors in indicating characteristic points on the body (regarding the joints and head). To fully analyze how people move, it is necessary to precisely find key points in the image of a man.

When expanding the system design, more advanced ways of analyzing human motility can be included in the future by placing precise GPS receivers or inertial sensors [6] on the human body and reading spatial coordinates from them. Another way may be to detect elements of movement more easily recognizable (such as a limp), rather than to increase the precision of image recognition in order to capture the smallest characteristics of human motion. Then such a system could be used to search for people who have had an accident, or have impaired mobility.

Studies have confirmed that human physical parameters have a great impact on distinguishing people from the environment. That is why, when designing this type of system, one should not limit oneself to characteristic motoric features, but use any information that can be extracted from the source data.

## REFERENCES

[1] Collective work, [Great PWN Encyclopedia], PWN, Warsaw (2008).
[2] "The Dangers of Facial Recognition Technology, " <https://blog.squareboat.com/the-dangers-of-facial-recognition-technology-97fa999319ac> (21 May 2019).
[3] Viola, P. and Jones, M., "Rapid Object Detection using a Boosted Cascade of Simple Features," Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 (2001).
[4] Andriluka M., Pishchulin L., Gehler P. and Schiele B., "2D Human Pose Estimation: New Benchmark and State of the Art Analysis" (2014).
[5] Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., and Dollár, P., "Microsoft COCO: Common Objects in Context," ECCV (2014).
[6] Paszek, J. and Kaniewski, P., "Simulation of Random Errors of Inertial Sensors," XIII International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science TCSET'2016, Lviv-Slavske, Ukraine, 23-26.02.2016, pp. 153-155 (2016).