PRINCIPLES OF ANALYSIS AND PROCESSING OF VIDEO INFORMATION IN THE IDENTIFICATION OF PERSONAL MOVEMENT

[1]Akhatov Akmal Rustamovich, [2]Himmatov Ibodilla Qudratovich

[1]Doctor of technical sciences, professor, vice-rector for international affairs of Samarkand State University named after Sharof Rashidov

[2]2st year PhD student of Samarkand State University named after Sharof Rashidov

https://doi.org/10.5281/zenodo.7952033

Abstract. *Intelligent analysis and processing of video data for personal identification is one of the promising, fastest growing research areas and is widely used in various fields of human activity. Initially, the information obtained from video cameras was mainly used in security television systems. This allows the use of new video data, the use of analysis and processing methods, as well as the recognition and identification of a person using his movements.*

Keywords: *identification, video image, gait recognition, movement, human pose estimation, data processing.*

Introduction

Today, great importance is attached to the development of the theoretical basis of image recognition and evaluation of human behavior. Therefore, it is represented by the presence of advanced mathematical apparatus methods in the characterization of objects in a complex state, in the recognition, intellectual analysis and processing of basic dynamic objects. Therefore, modern automated video information processing and analysis systems are widely used. Including solving problems such as video surveillance for security and access to objects for various purposes (security systems, control systems), fire protection systems, in navigation, for quality control and the quantity of manufactured products and video monitoring of many it is possible to use it in the control and identification of target processes. Unlike static frames, live video contains images and information about changes that occur during the observed time is stored in the state of the scene. Video sequences are a comprehensive information resource that combines the spatial properties of static images and their multidimensional properties are represented in the form of time-varying signals. Their main features include appearance, redundancy, compactness, maximum information content (up to 80% of all information delivered to a person).

Nowadays, the identification of a human by his actions is one of the processes that are difficult to set up and master compared to other identification technologies. Therefore, in this study, the recognition of human poses and the construction of graphs using neural networks and other tools are considered as initial tasks.

Human pose estimation is a special case of the image segmentation problem in the computer vision department, which consists in detecting the movements of the human body from parts of images or videos (considered as a sequence of images). Often a human's position is replete with associated key points that correspond to the joints (shoulders, elbows, arms, hips, knees, feet) and other key points (neck, head). This task can be considered in two or three dimensions, which determines the complexity of the task and the practical application of the results [1].

Literature review

Modern motion image recognition studies usually use established databases. Such data sets are mainly provided by researchers from Germany, Japan and India. The TUM Gait from Audio, Image and Depth (TUM-GAID) [2], the OU ISIR Large Population Dataset (OULP) [3] and the CASIA Gait dataset are the most widely used relatively perfect datasets for human motion detection. It is noteworthy that these collections are constantly updated and improved, ensuring that the information in the collections does not lose relevance.

The simplest database TUM-GAID is used to determine the side view of a person. It contains images for 305 cases from left-to-right and right-to-left side views. The frame rate of the videos is about 30 fps and the resolution of each frame is 320×240. The TUM-GAID kits contain videos that describe different walking conditions, namely 6 normal walking, 2 walking in loose clothing, and 2 walking in heels for each person. Although the images were recorded from only one view, the number of study subjects (people) allowed half of them to be taken as the sample training set and the other half as the evaluation set. The database is split into training, validation, and test sets, using data for 150 cases to train the models and another 155 for testing. Thus, no randomness affects the result. Being in full color, this set allows for many experiments with different approaches. In addition, this database contains images taken at six-month intervals, which allows to test the stability of the algorithms in the face of temporal gait changes.

The next dataset is CASIA B, which is relatively small in terms of the number of subjects, but it has a large variation of views, and it captures 124 subjects from 11 different viewing angles (0° from 18° to 180°) is obtained. Videos are 3 to 5 seconds long, have a frame rate of about 25 fps, and a resolution of 320×240 pixels. All trajectories were recorded directly and closed to all participants in the same room. In addition to the variety of solutions, as in the first collection, situations such as wearing a coat and carrying a bag are presented in different clothes. In total, there are 10 video sequences for each person taken from each view, including 6 simple walks without additional conditions, two walks in bare clothes, and two walks with a bag. The CASIA dataset, combined with its popularity due to its visual variability, has few cases to train a deep neural network that can recognize any type of walk without overfitting.

The methods of using deep neural networks were first studied as DeepPose by A. Toshev, in studies led by J. Thompson, an image of a map of hot spots of the main points of the body was used to improve their localization when a human moves a lot or regularly performs any activity [1,3]. The Markov Random Field (MRF) has promoted the use of a spatial model to evaluate the relationships of key points. H.Chu et al. developed a kernel transformation method to study the correlation between key points with a high degree of correlation using a bidirectional tree [3].

Discussion

The measurement of object parameters involves the estimation of the following defined parameters of video images [4]:

1) Determining and evaluating the speed of movement of objects in a video image (for example, measuring the speed of cars, identifying a person moving at a non-standard speed, that is, running, etc.);

2) calculating the number of objects of interest to the user: evaluating the intensity of traffic and counting the number of passing vehicles, automatically determining the number of busy and empty places, counting the number of products on the conveyor, conducting statistical studies in

shopping and entertainment centers, as well as , in museums, counting the number of people passing through a certain controlled area;

3) determining the dimensions of objects;

4) determining the distance from the object to the video camera.

Object recognition includes:

1) to recognize objects of a certain shape and size;

2) recognition of numbers and inscriptions on moving objects;

3) identification of people, including by biometric signs (face, gait, thermal portrait, etc.);

4) recognition of situations (early detection of fires, accidents, disasters, theft and non-standard behavior in the controlled area).

Within the framework of the above-mentioned tasks, the issue of identification of a person's movement is considered as the most urgent and complex problem.

As a result, the main goal of video analysis is to extract useful and relevant information from the frames in the sequence by monitoring. The importance of the criteria of this information is determined by the target tasks and the main task of a particular system.

Thus, it should be noted that automated video surveillance systems provide information about human actions, the situation formed in the field of vision of the camera, and information about making decisions presented to the software [5]. The process of building such systems is represented by the following scheme (Fig. 1).

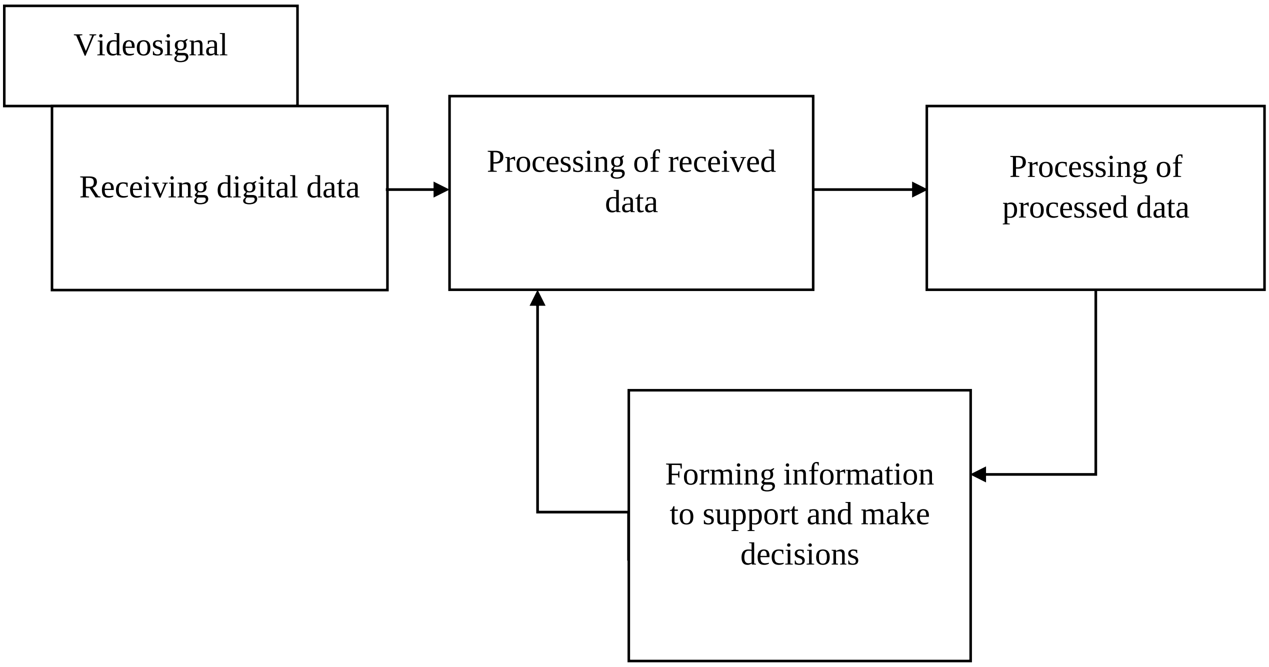
Videosignal
Receiving digital data
Processing of received data
Processing of processed data
Forming information to support and make decisions

Fig.1. Automated video surveillance system organization scheme

The above process includes the following steps:

1) receiving a video signal and converting it to digital form (if the main signal is received in analog form);

2) collecting current frames of video data;

3) processing of digitally received video images, highlighting relevant and necessary information;

4) analyzing the received data and choosing a specific task to solve it.

If the object to be recognized is viewed in relation to moving vehicles, the issue of person identification in video surveillance systems becomes more complicated [6]. Because in recognition, the video sequence is not represented by time passes, and video image transformations can be created based on certain analytical relationships [7].

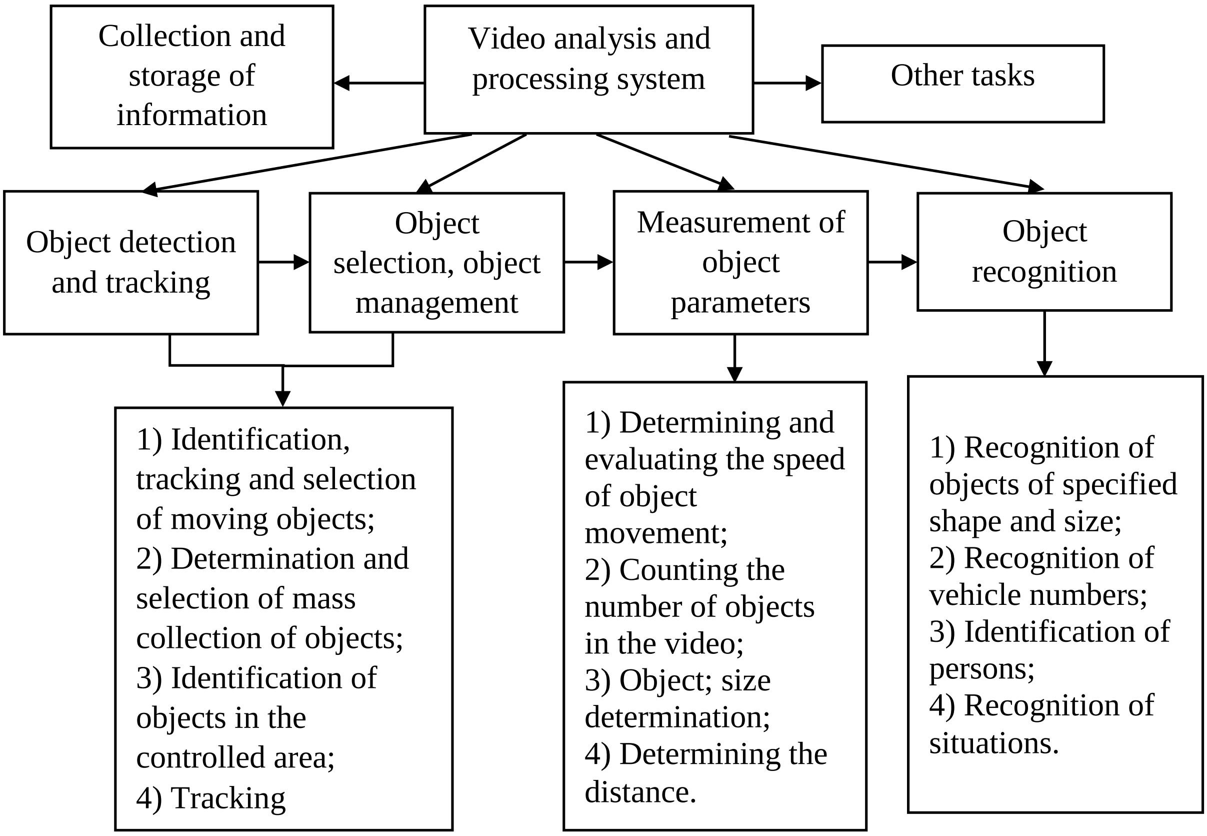
Collection and storage of information
Video analysis and processing system
Other tasks
Object detection and tracking
Object selection, object management
Measurement of object parameters
Object recognition
1) Identification, tracking and selection of moving objects;
2) Determination and selection of mass collection of objects;
3) Identification of objects in the controlled area;
4) Tracking
1) Determining and evaluating the speed of object movement;
2) Counting the number of objects in the video;
3) Object; size determination;
4) Determining the distance.
1) Recognition of objects of specified shape and size;
2) Recognition of vehicle numbers;
3) Identification of persons;
4) Recognition of situations.

Fig.2. Target tasks of image analysis and processing systems

It is appropriate to approach the tasks of classification, recognition and detection of motion as a single process in a single systematic position in the objects of the video sequence. Therefore, typical tasks of video image analysis (Fig.2) include [8]:

- identifying objects, selecting them (selecting the necessary information);
- data storage;
- identifying objects and tracking them;
- selection of moving objects and switching to the current management status;
- measurement of object parameters;
- classification and recognition of objects;
- other tasks (according to the specific characteristics of monitoring).

Image recognition problems lead to the following interrelated issues of selection and recognition [1, 3, 7]:

1) identification, selection and tracking of moving objects (people, animals, transport, etc.);

2) identification and selection of mass gathering of objects (crowded traffic, road traffic);

3) identification of objects located in the controlled area, control of protected objects (detection of the presence of objects abandoned by people in traffic areas, identification of improperly stopped vehicles);

4) determining and monitoring the trajectory of the movement of objects, i.e. identifying people moving in the wrong direction, turning in a prohibited direction, driving in the opposite direction, leaving the road are taken into account.

Similar to the problem of arbitrary computer vision, in motion detection there are many factors that do not change the content of the video in any problematic situation, but change the appearance and therefore the internal state of the gait. All these factors are divided into two groups: those affecting the image itself and those affecting only the condition. The factors of the first group really change the way you walk and are more difficult to recognize for both humans and computers. In the second group, there are conditions that change only the internal representation, and the way a person walks does not change. To the human eye, the two videos will show the same person, but to the computer, they will be two different image sequences with different characteristics.

Overall performance analysis shows that the hourglass can be improved by using multiple smaller filters instead of one larger filter, for example using two 3x3 filters instead of a 5x5 filter. In addition, the 1x1 filter to reduce the number of pixels with convolution improves its performance as well. Thus, this architecture uses full size filters of 3x3 or less. In addition, it should be noted that 64x64 input images are submitted to the network instead of high resolution images to avoid excessive use of GPU memory. This does not negatively impact performance.

Results

These methods and principles are the mechanism, the problem of calculating the average frame characteristics of all mentioned methods leads to a partial loss of temporal information, because the order of the frames is not taken into account.

Several approaches aim to prevent this loss [7, 8]. Thus, the Chrono-Gait Images (CGI) descriptor extracts contours from silhouettes and encodes temporal information based on the position of the corresponding frame in the gait cycle [9]. Another gait descriptor is the Gait Flow Image (GFI), which is based on the binarized optical flow between successive silhouette images and represents the gradual change of silhouettes [10]. Another way to describe silhouette-based gait that preserves temporal information is to determine gait entropy in the frequency domain [11].

This approach combines the walk entropy identification method with Discrete Fourier Transform (DFT), which averages binary silhouette masks with exponentially decreasing amounts. The walk entropy method in the frequency domain is an integration of approaches that calculates the entropy based on the discrete Fourier transform rather than the walk energy image [12].

Conclusion

Thus, the article considers different approaches to the combination of all the mentioned factors, the problem of recognition becomes more complicated, in some cases outerwear, conditions of cargo transportation, height of shoe heels are also taken into account.

Due to the extreme variability of the influencing conditions, it is difficult to form a large-scale general data set. Usually, researchers focus on certain conditions and ignore other conditions, which in turn leads to overfitting of models to specific conditions presented in the database. It should be noted that no matter how large the database is, it remains impossible to cover all conditions, and this situation justifies the relevance and necessity of the research topic of this dissertation.

The use of deep analysis algorithms proposed in this study using neural networks allows fast processing of these captured images. In the future, this will make it possible to determine the psychological characteristics of a human using algorithms and neural networks to form behavior and express its psychological nature.

REFERENCES

1. Akhatov A., Nazarov F. Rashidov A. Increasing data reliability by using bigdata parallelization mechanisms. International conference on information science and communications technologies. 4,5,6 November. ICISCT 2021(IEEE), art. no. 9670387 DOI: 10.1109/ICISCT52966.2021.9670387 SCOPUS.
2. Ximmatov I.Q. Advantages of biometrik gait recognition. Important factors in evaluation of gait analysis systems. SamDU Ilmiyi axborotnomasi. ISSN 2091-5446, 2020-yil, 3-son (121), [104-107].b
3. Ximmatov I.Q. Important factors in evaluation of gait analysis systems and advantages of biometric gait recognition. Innovatsion va zamonaviy axborot texnologiyalarini ta’lim, fan va boshqaruv sohalarida qo‘llash istiqbollari.Samarqand, 14-15 may, 2020 y. [262-267].b
4. Axatov A.R, Ximmatov I.Q. Foydalanuvchilarni biometrik autentifikatsiya turlari asosida haqiqiyligini tasdiqlash usullarinning samaradorligi. Innovatsion yondashuvlar ilm-fan taraqqiyoti kaliti sifatida: yechimlar va istiqbollari. Jizzax, 8-10 oktyabr 2020 y. [20-26].b
5. Alejandro Newell, Kaiyu Yang, and Jia Deng. Stacked Hourglass Networks for Human Pose Estimation// 26 Jul 2016 P. 3-4.
6. Jinbao Wang, Shujie Tan, Xiantong Zhen, Feng Zheng, Zhenyu He, Ling Shao “Deep 3D human pose estimation: A review” Journal of “Computer Vision and Image Understanding” Volume 210, September 2021, 103225.
7. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D human pose estimation: New benchmark and state of the art analysis. In: CVPR. pp. 3686–3693 (2014)
8. Bourdev, L., Malik, J.: Poselets: Body part detectors trained using 3D human pose annotations. In: ICCV. pp. 1365–1372 (2009)
9. Bulat, A., Tzimiropoulos, G.: Human pose estimation via convolutional part heatmap regression. ECCV pp. 717–732 (2016)
10. Chang, M., H.Qi, Wang, X., Cheng, H., Lyu, S.: Fast online upper body pose estimation from video. In: BMVC. pp. 104.1–104.12. Swansea, England (2015)
11. Cherian, A., Mairal, J., Alahari, K., Schmid, C.: Mixing body-part sequences for human pose estimation. In: CVPR. pp. 2361–2368 (2014)
12. Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A.L., Wang, X.: Multi-context attention for human pose estimation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5669–5678 (2017)