Development of neural network-based consultant recognition method for determining posture and behavior

Rozaliev V.L., Alekseev A.V., Ulyev A.D., Orlova Y.A.
Volgograd State Technical University
Volgograd, Russia
vladimir.rozaliev@gmail.com, alekseev.yeskela@gmail.com,
ulyev-ad@yandex.ru, yulia.orlova@gmail.com

Petrovsky A.B., Zaboleeva-Zotova A.V.
Federal Research Center
Computer Sciences and Control, RAS
Moscow, Russia
pab@isa.ru, zabzot@gmail.com

***Abstract*—The article presents the method recognition of sales consultants on the basis of a neural network to determine the posture and clarifying algorithms, also methods of monitoring the behavior of the seller-consultant and analysis of its interaction with the buyer. A brief review of the analog systems is given. The description of the proposed method is presented, the obtained results and ways of improvement are shown.**

***Keywords*—neural network, artificial intelligence, recognition of human pose, behavior monitoring.**

I. Introduction

The modern era is characterized by a transition from the economy of producers to the economy of consumers. In the conditions of toughening competition in the sphere of trade and rendering services, client-oriented services acquire special importance.

The main problem of introducing such services is the human factor, control of which is problematic due to the lack of ready-made software products.

Ensuring the proper quality of service delivery becomes the main objective of the market strategy for business development.

To improve the quality of service, it is proposed to develop and implement a software product to monitor the activities of consultant salesmen through the analysis of their work with the use of equipment for video fixing [10].

The work of the software product is based on the algorithm Pose Estimator [3], which allows you to determine the position of a person, clarifying algorithms, auxiliary neural networks that help to identify the seller-consultant, and also determine the quality of services provided to them [2].

II. The proposed methodology

To solve the problem, we propose to use a cascade of two neural networks:

- Fast convolutional neural network FastPoseEstimator, trained on mobilenet architecture;
- Algorithm of stabilization of "key" points, allowing determine those points of the body, which could not recognize the neural network;
- Neural network for determine the behavior of the sales assistant and store employee.

III. The first stage, the use of the neural network "Fast Pose Estimator".

The main task of the neural network is to establish a person's pose through a nonparametric representation called the Part Affinity Fields (PAFs) by developers, to further determine the location of the seller's uniform of the consultant (branded T-shirt, cap, etc.).

The main advantage of the neural network is the high speed of the work, 1 frame in 1-2 seconds on GPU and 5-8 seconds on CPU. The main drawback is the decrease in the quality of work when compared with the classic version of Pose Estimtor.

Input data for the algorithm "PoseEstimator" is a graphic image of the sales consultant, on the output - an image with the selected parts of the human body.

The result of this network can be seen in figure 1.

Figure 1. The result of the Fast Pose Estimator neural network.

IV. The second phase, a stabilization algorithm "key" points of the a person body.

The skeleton is built from the starting point located under the throat, then the eyes, shoulders and pelvis are recognized. From the shoulders completed the hands, from the points of the pelvis of the foot, from the eyes ears. Thus, the key point for the construction is the point under the neck, then we call it the initial one.

It is necessary to determine the dominant color of the t-shirt seller. To improve the quality of work, the following algorithm is proposed for finding the missing "key" points, based on information about the structure of the human structure. Consider one of the options for finding the missing "key" point of the body.

If only one shoulder of a person is recognized, the second shoulder can be completed by depositing an equal segment from the projection of the starting point on the normal of the point of the found shoulder to the side of the not found shoulder equal to the distance from the projection of the point to the normal of the point of the found shoulder to the point of the found shoulder.

Thus, using the information about the structure of the human structure it is possible to determine the approximate location of the missing points of the shoulders, pelvis points. If there is only a starting point, it is proposed to determine some area of a fixed size on the human's body by depositing the area below the key point by N pixels. An example of the algorithm is shown in figure 2.

Figure 2. Example of detection of the second shoulder.

V. The third stage, the determination of the dominant colour in the uniform section.

The main task of this stage is to establish a dominant colour in the area of the uniform of a person to determine it in the group of sales consultants. Within the framework of this algorithm, an image from the "Pose Estimator" with the tops of the human body parts is input. On the basis of which there is a selection of the necessary clothing of a person [4] [5] [6]. To implement the definition of dominant colour in an established area, there are several methods: determining the ratio of a pixel to a given set of colours and clustering by the k-means method.

In the first method, the image is converted to HSV colour space, after which all pixels of the image are analyzed and based on the Hue, Saturation, Value data, the colour is set.

The idea of the k-means method [7] is to minimize the total quadratic deviation of the cluster points from the centers. At the first stage, you select points (three-dimensional RGB space) and determine whether each point belongs to this or that center. Then at each stage, the centers are redefined until a single center is found.

VI. Training the appearance of sales consultants.

To determine only the colour of a person's clothing is not enough to classify him as a sales consultant group. It is necessary to take into account the conditions of the difference in the illumination of the room at different times of the day, as well as the likelihood that there may be clarified and dark areas in the room. Thus, the recognized colour of the shape can vary.

To solve this problem, it is necessary to teach the system all possible colors that can be "read" from the clothing of the seller-consultant.

The operator of the software product at the beginning of work with the program should start the training mode, with by means of which prevailing color uniforms randomly moving around the store sales assistant.

On the basis of the prevailing colors of the clothes is formed an "average dominant color", which is the central point in the formation of the color range for referring a person to a sales consultant [1].

VII. Analysis of the quality of communication with the seller by buyers.

After the classification of people in the frame to groups of buyers and sellers, the program automatically controls the quality of services rendered by sellers-consultants. To assess the quality of the seller's communication with the client, it is proposed to use a set of algorithms and a cascade of neural networks.

The main task of the algorithms is to determine the location of the seller next to the buyer, as well as control over the personalization of the employee's appeal to the buyer.

In order to exercise control over the personalization of the employee's appeal to the buyer, it is proposed to determine the "field of view" of people in the frame. The seller must always interact with the buyer, be in the "review" area of the buyer and talk about the benefits of the goods. Thus, to accomplish the task, it is necessary to build a "field of view" based on the location of the eyes and ears obtained from Pose Estimator, after which the sector of intersection of these areas and the "angle

of interaction" between the buyer and the seller should be determined.

If the areas do not intersect or the "interaction angle" is less than the angle set by the operator of the software product, it is considered that the seller is near the buyer, but does not interact with it. Examples of definition algorithms for the "field of view" are presented in figure 3.

To determine the interest and emotionality of the store employee when interacting with the buyer, it is proposed to use neural networks that define the above parameters by the store employee behavior: hand speeds, moving around the store to accompany the buyer, movements to show goods, etc.

In addition, it is planned to use a separate neural network, which, by facial expression and behavior, will determine customer satisfaction with the services provided by the seller's. The conclusion about the work seller's of the store is made on the basis of all the factors that determine the interaction with the buyer, after which at the end of the month an estimate is calculated for each store employee.

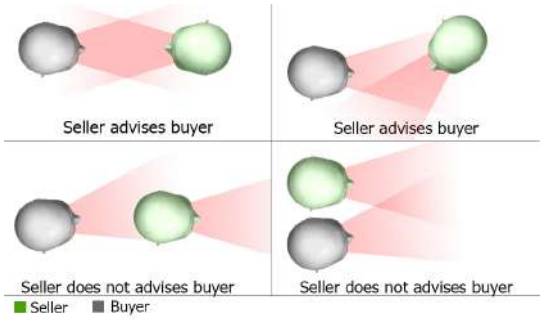

Figure 3. An example of the definition of "field of view".

VIII. Overview of analogues.

It should be noted that the finished software products that allow to solve the problem discussed in this article are not present. Similar software products perform only part of tasks.

The simplest example of intelligent video surveillance is motion detection. One detector can replace several video surveillance operators. And in the 2000s, the first video analytics systems began to appear, capable of recognizing objects and events in the frame. Most of the solutions work with face recognition technologies. Solutions in this area include Apple, Facebook, Google, Intel, Microsoft and other technology giants. Surveillance systems with automatic passenger identification are installed in 22 US airports. In Australia, they are developing a biometric system of face recognition and fingerprinting within a program designed to automate passport and customs control. An interesting project of NTechLab company showed a system capable of real-time recognition of sex, age and emotions using the image from a video camera. The system is able to evaluate the audience's reaction in real time, so you can identify the emotions that visitors experience during presentations or broadcasts of advertising messages. All NTechLab projects are built on self-learning neural networks. In our system, we do not yet use data on a person's face. We plan to process this information at the next stages of the project development.

In other systems, the object tracking function is used - tracking. The operation of the tracking modules is related to the operation of the motion detector. To construct the trajectories of the movement, a sequential analysis of each frame is carried out, on which moving objects are present. The simplest implementation of tracking considers two frames and builds trajectories along them. First, the movements on the current and previous frame are marked, then, by analyzing the speed, the direction of movement of objects, and also their sizes, the probabilities of the transition of objects from one point of the trajectory of the previous frame to another point of the current are calculated. The most probable movements are assigned to each object and added to the trajectory. Objects in the frame can move in different ways: their trajectories may intersect, they can disappear and arise again. To improve the accuracy of tracking, some manufacturers use the technology of sequence analysis and continuous post-processing of the results obtained. We have planned to use this approach in our system.

Another analogue of our system - GPS-trackers. These systems work based on the definition of geolocation. To implement this solution, each employee must be equipped with a separate GPS tracker, the data from which will be sent to the server at some interval. However, this solution has a number of drawbacks:

1) The solution is not cost-effective, since it is necessary to purchase GPS trackers for all personnel.
2) We can't exclude the situation in which the seller can give his GPS-tracker to a partner to deceive the system.
3) Such a solution is not universal. When identifying sales consultants through the camera, it is possible to expand the functionality, determine the level and time of interaction of the seller with the buyer, and much more.

Also, analogs include systems for counting the number of visitors on a video stream. These systems also have a number of shortcomings, the main one of which is the impossibility of identifying sales consultants and the quality of their services.

IX. Conclusion.

In order to improve the technological process of detecting the seller's consultant, it is possible to develop

additional functionality.

To more accurately determine the seller's consultant, it is possible to analyze several elements of the uniform at once (for example, a yellow T-shirt and black pants).

Another factor that allows to detect the seller, can serve as a definition of behavior, characteristic for the seller-consultant. To solve this problem, you will need to create another neural network.

It is also possible to identify additional factors that determine the quality of the seller's work in the store, such as the presentation goods to customers who have not stopped at the rack with the goods, but passing by.

In addition, it is planned to implement the ability to install "dead zones" in the program. This function allows you to set the "non-human" type for those objects that the Fast Pose Estimator neural network defines as people.

Thus, the developed software will make it possible to qualitatively improve the work of the sales assistant and, as a result, will lead to an improvement in the customer focus of the business. Figure 4 shows an example of the program.

This work is a contin of the work, where the features and possibilities of determining the post-sense of its semantic distinctive feature were considered [8] [9]. This work was partially supported by RFBR (grants 17-29-07021, 18-47-340006, 18-47-342002, 18-07-00220, 19-07-00020).

Figure 4. An example of the program when recognizing an employee of the store.

REFERENCES

[1] M.D.Khorunsjiy Metod kolichestvennoi otsenki tsvetovykh razlichii v vospriyatii tsifrovykh izobrazhenii. [The method of quantitative estimation of color differences in the perception of digital images.]. Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki. [Scientific and technical herald of information technologies, mechanics and optics], 2008

[2] O. Ulyanova Psihologicheskie osobennosti prodavcov-konsul'tantov setevogo marketinga. [Psychological features of network marketing sales consultants.]. 2013, Retrieved from https://cyberleninka.ru/article/n/psihologicheskie-osobennosti-prodavtsov-konsultantov-setevogo-marketinga

[3] Cao Z., Simon T., Wei S.-E., Sheikh Y. Otsenka pozy v real'nom vremeni Multi-persony 2D s ispol'zovaniem polei blizosti. [Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.]. 2016, Retrieved from http://arxiv.org/abs/1611.08050

[4] Zhe Cao Otsenka 2D-otsenki v real'nom vremeni s ispol'zovaniem otdel'nykh polei. [Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.]. 2017, Retrieved from http://www.ri.cmu.edu/wp-content/uploads/2017/04/thesis.pdf

[5] U. Iqbal, J. Gall Otsenka individual'nosti cheloveka s uchastiem mestnykh assotsiatsii. [Multi-person pose estimation with local joint-to-person associations.]. 2016, Retrieved from https://arxiv.org/pdf/1608.08526.pdf

[6] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, B. Schiele Bolee glubokii srez: bolee glubokaya, bolee sil'naya i bolee bystraya model' otsenki pozy dlya neskol'kikh chelovek. [Deepercut: A deeper, stronger, and faster multi-person pose estimation model.]. 2016, Retrieved from https://arxiv.org/pdf/1605.03170.pdf

[7] Y. Osipove, D. Lavrov Primenenie klasternogo analiza metodom k-srednih dlya klassifikacii tekstov nauchnoj napravlennosti. [Application of cluster analysis by k-means method for classification of scientific texts.]. 2017, Retrieved from https://cyberleninka.ru/article/n/primenenie-klasternogo-analiza-metodom-k-srednih-dlya-klassifikatsii-tekstov-nauchnoy-napravlennosti.

[8] V.L.Rozaliev, Y.A.Orlova Opredelenie dvizhenii i pozy dlya identifikatsii emotsional'nykh reaktsii cheloveka. [Recognition of gesture and poses for the definition of human emotions]. 11-ya Mezhdunarodnaya konferentsiya po raspoznavaniyu obrazov i analizu izobrazhenii: novye informatsionnye tekhnologii (PRIA-11-2013), Samara, 23-28 sentyabrya 2013 g.: Trudy konferentsii [11th International Conference of Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013), Samara, September 23-28, 2013 : Conference Proceedings], 2013, vol. 2, pp. 713-716

[9] A.S.Bobkov, V.L.Rozaliev Fazzifikatsiya dannykh, opisyvayushchikh dvizhenie cheloveka. [Fuzzification of data describing the movement of a person]. Otkrytye semanticheskie tekhnologii dlya proektirovaniya intellektual'nykh sistem (OSTIS-2011): mater. stazher. nauchno-tekhnich. konf. (Minsk, 10-12 fevralya 2011 g.) [Open semantic technologies for the design of intelligent systems (OSTIS-2011) : mater. intern. scientific-techn. conf. (Minsk, Feb. 10-12. 2011)], 2011, pp. 483-486/

[10] Angjoo Kanazawa, Michael J. Black, David W. Jacobs, Jitendra Malik Kompleksnoe vosstanovlenie formy cheloveka i pozy [End-to-end Recovery of Human Shape and Pose]. Retrieved from https://www.researchgate.net/publication/321902575_End-to-end_Recovery_of_Human_Shape_and_Pose?discoverMore=1

РАЗРАБОТКА МЕТОДА РАСПОЗНАВАНИЯ ПРОДАВЦОВ-КОНСУЛЬТАНТОВ НА ОСНОВЕ НЕЙРОСЕТИ ДЛЯ ОПРЕДЕЛЕНИЯ ПОЗЫ И ПОВЕДЕНИЯ

Розалиев В.Л., Алексеев А.В.,
Ульев А.Д.,Орлова Ю.А.,
Петровский А.Б., Заболеева-Зотова А.В.

В статье представлен метод распознавания продавцов-консультантов на основе нейросети для определения позы и уточняющих алгоритмов, а также рассмотрены методы контроля поведения продавца-консультанта и анализа его взаимодействия с покупателем. Проведен краткий обзор систем с похожими функциональными характеристиками. Представлено описание предлагаемой методики, показаны полученные результаты и пути улучшения

Received 10.01.19