Automatic recognition of consultants on video records

Rozaliev V.L.
CAD Department
Volgograd State Technical University
Volgograd, Russia
vladimir.rozaliev@gmail.com

Alekseev A.V.
CAD Department
Volgograd State Technical University
Volgograd, Russia
alekseev.yeskela@gmail.com

Ulyev A.D.
CAD Department
Volgograd State Technical University
Volgograd, Russia
ulyev-ad@yandex.ru

Orlova Y.A.
SAS Department
Volgograd State Technical University
Volgograd, Russia
yulia.orlova@gmail.com

Petrovsky A.B.
Laboratory of methods and decision support systems
Federal Research Center Computer Sciences and Control, RAS
Moscow, Russia
pab@isa.ru

Zaboleeva-Zotova A.V.
Laboratory of methods and decision support systems
Federal Research Center Computer Sciences and Control, RAS
Moscow, Russia
zabzot@gmail.com

***Abstract*—This article presents method for recognizing consultants in showroom based on cascade of neural networks. The cascade consists of two networks - high-performance detector and refining module with recognition of pose. 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, analysis of video stream**

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 [11].

The basic principle of the software product is based on a neural network for detecting a person on a frame from a video stream, and also on the algorithm "Pose Estimation" [3], whose main function is to recognize the human pose.

II. The proposed methodology

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

- fast detector Yolo [1];
- Neural network for recognition of the pose [2] of the consultant, the general scheme is shown in figure 1.

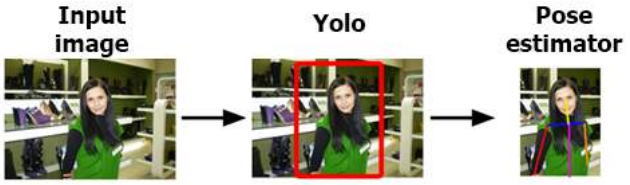

Figure 1. The proposed method for recognizing consultants.

III. The first stage, the use of the "YOLO" neural network

The first stage of the cascade is the neural network [1]. The main advantage of Yolo is the speed of work, it allows you to achieve a speed of 60-90 frames per second, while maintaining a sufficiently high quality of work. The input data of the algorithm "Yolo" is the image, and the output data - rectangles (wireframe figures) that limit the found objects (this network can find other objects, but we are only interested in people). These values (regions) are transmitted to the second neural network for clarification.

The main task solved by the Yolo network in our approach is a quick determination of the fact of having people in the frame, so as not to start a slower stage on empty frames. Advantage in the speed of work is achieved due to the approach classifiers that are different from the classical networks. In the case of

Yolo, a picture is placed at the entrance, divided into small regions with the probabilities of finding objects in the region. The result of the network is shown in figure 2.

Figure 2. The result of the work of the Yolo neural network.

IV. The second stage, the use of the neural network "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.) [4] [5] [6].

The main advantage of the neural network is the high quality of the work. The main drawback is the demanding nature of the neural network "PoseEstimator" for the technical characteristics of the equipment used. In the absence of an appropriate technical base for this network, there is a sharp increase in the processing time.

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 3.

Figure 3. The result of the Pose Estimator neural network.

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. An example is shown at figure 4.

Figure 4. Allocation of the seller's clothing segment.

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 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. An example of clustering is shown at figure 5 [7].

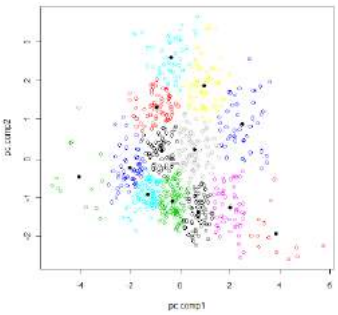

Figure 5. Example of clustering.

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 [10].

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 start of work with the program must manually select on

the frame of the seller-consultant, which will arbitrarily move around the room.

In addition, to determine the degree of colour deviation from the established colour of the seller-consultant, the colour difference formula is used, shown at formula 1, which will allow numerically to express the difference between the two colours in colorimetry. On the basis of the data obtained, it can be concluded on how much the colour is "close" to the colour established for the seller-consultant.

$$\Delta E^*_{ab} = \sqrt{(L^*_2 - L^*_1)^2 + (a^*_2 - a^*_1)^2 + (b^*_2 - b^*_1)^2} \qquad (1)$$

VII. 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 [12].

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. In the general case, several moving objects can be present in one frame, so the program needs not only to construct trajectories, but also to distinguish objects and their movements. 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. The program builds graphs - it analyzes the transitions of objects from one state to another. In order to understand which object the movement corresponds to, the speeds and directions of motion, position, color characteristics are also analyzed. As a result, a set of the most probable displacements of the object is formed, forming a trajectory. 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. An example of the work of such products is shown at figure 6.

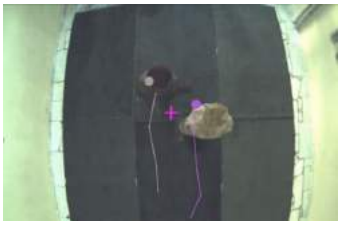

Figure 6. Example of a program for counting the number of visitors.

VIII. 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).

In addition, it is possible to search for the company logo on the uniform, the location of which will allow us to identify with confidence the person as the seller-consultant.

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.

Thus, the developed software product, consisting of a cascade of neural networks YOLO and Pose Estimator will allow to qualitatively improve the work of the seller-consultant and, as a result, improve the client-oriented business.

Figure 7 shows the input image for the system, in figure 8 - the obtained result.

This work is a contin of the work [8] [9], where the features and possibilities of determining the post-sense of its semantic distinctive feature were considered.

This work was partially supported by RFBR (grants 16-07-00407, 16-47-340320, 16-07-00453, 18-07-00220).

Figure 7. Image to be input.

Figure 8. The image obtained as a result of the software product.

REFERENCES

[1] Redmon, J., Farhadi A. YOLO9000: luchshe, bystree, sil'nee [YOLO9000: Better, Faster, Stronger]. Retrieved from http://arxiv.org/abs/1612.08242

[2] A.A. Shpirko, V.L. Rozaliev, J.A. Orlova, A.V. Alekseev Avtomatizatsiya postroeniya vektornoi modeli tela cheloveka [Automation of constructing a vector model of a human body.]. Izvestiya VolgGTU. Seriya "Aktual'nye problemy upravleniya, vychislitel'noi tekhniki i informatiki v tekhnicheskikh sistemakh" [Izvestiya VSTU. A series "Actual problems of control, computer technology and informatics in technical systems"], 2013, vol. 7, pp. 67-71/

[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] M.S.Shmakov, A.A. Tovmasyan Analiz tsvetovoi palitry izobrazhenii dlya opredeleniya preobladayushchikh tsvetovykh tonov [Analysis of the color palette of images to determine the prevailing color tones]. Raboty Belorusskogo gosudarstvennogo tekhnicheskogo universiteta. [Works of the Belarusian State Technical University], 2016, vol. 9

[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] 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

[11] 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

[12] Korobkov A. Otslezhivanie ob'ektov v videopotoke. Metody postroeniya traektorii. [Tracking objects in the video stream. Methods for plotting trajectorie]. Sistemy bezopasnosti. [Security], 2014, val. 3

АВТОМАТИЧЕСКОЕ РАСПОЗНАВАНИЕ КОНСУЛЬТАНТОВ НА ВИДЕОЗАПИСИ

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

В данной статье представлен метод распознавания консультантов в торговом зале на основе каскада нейронных сетей. Каскад состоит из двух сетей - высокопроизводительного детектора и уточняющего модуля с распознаванием позы. Представлено описание предлагаемой методики, показаны полученные результаты и пути улучшения.