Customer Traffic Distribution Analysis Based on Video Information

Tatyana Martynenko[a], Tatyana Vasyaeva[a], Aida Velieva[a] and Yuriy Skobtsov[b]

[a] *Donetsk National Technical University, Donetsk, Ukraine*
[b] *Saint Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia*

Abstract

The map constitution task of the customer movement through the store using videoanalitics has been considered. The problem is reduced to the video stream objects detection with further tracking. It is proposed to use pre-trained CNN for object detection. The experimentally justified joint use of pre-trained networks MobileNet-SSD. For the object tracking there have been performed experiments with algorithms built into the OpenCV library: GOTURN, CSRT, KCF, BOOST, TLD, MOSSE, MedianFlow, and MIL. According to the multiple object tracking accuracy (MOTA), the MedianFlow tracker is selected. Experiments were performed using a set of video sequences containing various negative parameters confirmed the effectiveness of the selected solutions.

Keywords 1

Computer vision, video analytics, deep learning, convolutional neural networks, detection, customer flow, tracking, conversion rate

1. Introduction

Today the retail sector is experiencing fierce competition and an increase in consumer demands for service levels. To maximize the effectiveness of marketing and sales an impressive array of IT solutions have been offered. One of them is video analytics [1]. In the retail sector video analytics provides significant competitive advantages, it allows you to evaluate such important parameters as the number of visitors, conversion rate (paying attention to a particular product). You can use it to get useful information about your customers and use it in the future to stimulate customer activity as well as to optimize the trading process through timely and effective personnel management.

Video analytics using computer vision methods can produce continuous automated data collection, analyzing the sequence of images coming from video cameras in real time or from archival records without additional staff.

According to undertaken studies [2] video analysis and computer vision technologies reduce by 10% the number of people leaving the store without buying something and by 20% the loss of store profits, and sales of individual products can be increased by 15-25% when changing their location according to the detected “hot zones”.

Thuswise, at the moment one of the most promising areas for analyzing customers behavior in a retail store is technology based on video analysis. It allows you to determine customer traffic statistics quickly and effectively, to create a portrait of the target audience, and to study the customer activity.

Russian Advances in Artificial Intelligence: selected contributions to the Russian Conference on Artificial intelligence (RCAI 2020), October 10-16, 2020, Moscow, Russia
EMAIL: tatyana.v.martynenko@gmail.com (T. Martynenko); vasyaeva@gmail.com (T. Vasyaeva); velievaaida9@gmail.com (A. Velieva); ya_skobtsov@list.ru (Y. Skobtsov)
ORCID: 0000-0002-1483-8483 (T. Martynenko); 0000-0001-9362-2279 (T. Vasyaeva); 0000-0002-5362-6256 (A. Velieva); 0000-0002-7677-2010 (Y. Skobtsov)

2. Articulation of an issue

The basic aim of any trading network is to get the maximum profit. This is achieved by increasing sales (due to an increase in the number of customers) and reducing costs (including reducing the number of staff avoiding the service degradation).

One of the key terms of video analytics in the retail industry is the customer traffic. According to [3], the customer traffic (customer flow) is the direction that most customers in the store follow.

The owner of a point of sale needs to have an up-to-date idea of the institution's attendance, about the movements of visitors inside the sale area, since this information is used to build a strategy for attracting and retaining customers, which in its turn is based on:

- optimization of the work of staff (to adjust the number of staff according to the intensity of the customer traffic in different periods of time);
- sales conversions [4] for selected departments or the store as a whole. Conversions show the ratio of the number of visitors to the point of sale in relation to the number of transactions (purchases);
- increasing the growth of sales of unpopular products due to their placement in the so-called "hot zones", i.e. the most popular places to visit in this point of sale;
- successful placement of advertisements and promotions in the departments that attract the most interest from visitors;
- changing the product layout based on the map of customers movements in the shop.

The main source of data for analysis is video cameras located above the entrance and exit, departments of supermarket and cash registers. A generalized plan of sales area is shown in Fig. 1.

The use of video analytics involves automation of four main functions [1]: detection, tracking, recognition, forecasting.

Tracking via detection is used to analyze the distribution of customer traffic [5]. This approach makes it possible to use high-precision object detection methods without a large computational load of the system due to tracking of already detected objects, excluding their repeated detections.

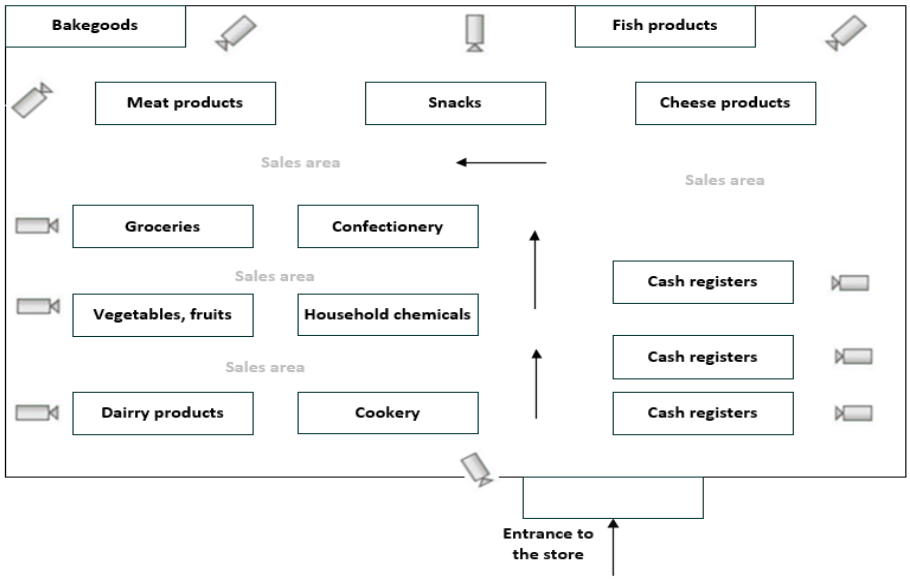

Figure 1: Plan of the sales area with video cameras.

The study object of this research is the process of detecting and tracking customers based on video information.

The aim of this work is to constitution a map of the customer movement around the store based on video information due to the usage of modern methods for detecting and tracking, which will make it possible to make effective managerial decisions in the retail industry.

3. Research Objective

3.1. Detection of Video Sequence Objects

The task of detecting in a video stream should be understood as detecting pre-defined classes of objects (people, vehicles, furniture, animals and so on) with determining the label and coordinates of the object's location [6].

You can represent the object`s location in different ways such as the set of pixels that correspond to the object [7] or the coordinates of a rectangle that bounds the object [8]. In this research we will get dozens of bounding rectangles (bounding boxes) at the output of the detection algorithm.

The input information of the developed system is the video stream S, which is made as a sequence of frames $I_1, I_2, \dots, I_k, \dots, I_N$.

$$I_k = \{I_k(x,y), 0 \le x < width, 0 \le y < height\}, k = \overline{1,N}, \quad (1)$$

where *width* – is the width of frame, *height* – is the height of frame, $I_k(x,y)$ – is the feature vector of colors, N – is the number of frames; k – is the frame number.

The set containing classes $C = \{c_1, c_2, \dots, c_i, \dots, c_M\}$. Our task is to detect objects $X = \{x_1, x_2, \dots, x_c\}$, with their subsequent selection by belonging to a given class (people's figures):

$$P_{c_i} = \{P_{c_i,1}, P_{c_i,2}, \dots, P_{c_i,n_c}\} \in X, \quad (2)$$

where P_{c_i} – is the set of detected objects belonging to the class c_i; n_c – is the number of detections.

The bounding box, which characterizing the location of the object in the image can be represented by the formula:

$$\{Out_k | I_k \in S\} Out_k = \{P_i\}, i = \overline{1,n_k}, \quad (3)$$

where Out_k – bounding box, n_k – the number of selected objects on a k-frame.

In such a case at the output of the detection algorithm one frame is made asset of dozens bounding rectangles that border objects of interest (people figures):

In such a case at the output of the detection algorithm we have many bounding boxes corresponding to objects of interest on one frame:

$$I_k = \{Out_1(x,y,w,h,c), Out_2(x,y,w,h,c) \dots Out_n(x,y,w,h,c)\}, \quad (4)$$

where (x,y) – are coordinates for each object in the image; (w, h) – is the dimensions of the object, given the width (w) and height (h); c – is the class connected with each bounding box.

3.2. Tracking of Video Sequence Objects

Tracking of moving objects (tracking) is the creation of a trajectory of movement of target objects in time by localizing its position on the input sequence of frames.

An object's movement trajectory is a sequence of its positions:

$$T = \{Out_s(x,y,w,h,c), Out_{s+1}(x,y,w,h,c), \dots, Out_{s+l-1}(x,y,w,h,c)\}, \quad (5)$$

where s – is the number of first frame in which the object was detected, l – is the number of frames sequence in which the object is observed, x and y – are coordinates of location; w – is the width; h – is the height and c – is the class number of the object in the video image.

To evaluate tracker accuracy, the MOTA criterion is typically used [9]. Mathematically, the MOTA criterion is described by the following formula:

$$MOTA = 1 - \frac{\sum_t m_t + fp_t + mme_t}{\sum_t g_t}, MOTA \to max, \quad (6)$$

where m_t – is the number of misses upon detection (false detection); fp_t – is the number of false positives; mme_t – is the number of mismatches; g_t – is the number of people present at time t.

4. Analysis of the Convolutional Neural Networks Usages in Object Detection Tasks

Detection is fundamental and one of the most difficult tasks of computer vision. Deep learning methods [6], in particular artificial neural networks [10], have become a powerful tool to solving them.

Algorithms based on convolutional neural networks (CNN) show the best quality in object detection tasks. CNN is a special neural network architecture proposed by Yann LeCun which is the main one used in computer vision [11]. A distinctive feature of CNN is the detection of objects in video images with an accuracy that exceeds the accuracy of other video image detection methods.

The classical CNN [11] has a hierarchical architecture (Fig. 2), and usually includes: convolution layers (convolution layer), pooling layers (pooling layer), and fully connected layers (dense layer).

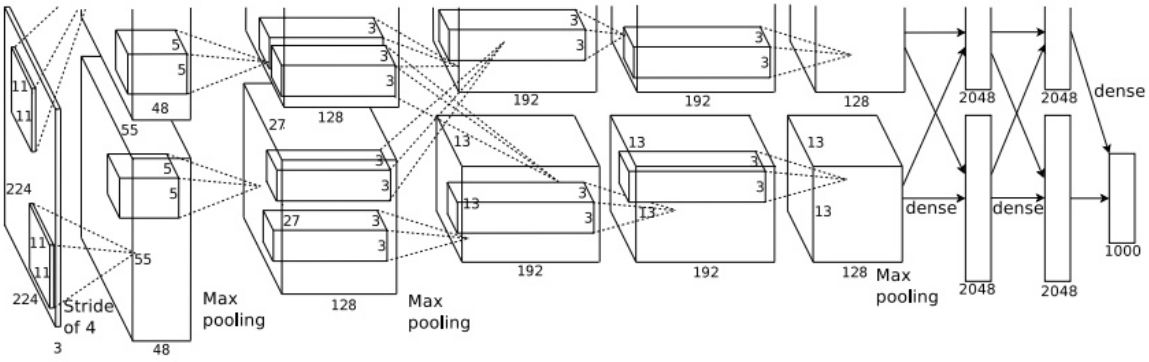

Figure 2: The classical CNN architecture

The differences between algorithms based on CNN use are in the configuration and selection: parameters of architecture (the number and type of layers, the value of weights, the number of neurons on each layer); training parameters; data that the neural network is trained on, as well as methods for input features processing.

Nowadays there are a large number of pre-trained convolutional networks, i.e. those that are already trained on a large data set within the large-scale task of video images detecting.

Detectors based on the use of convolutional neural networks can be divided into several groups: two-stage, one-stage, and basic algorithms.

The idea of two-stage algorithms (Region Proposal Networks, RPN) is to initially search for sets of sites, each of which assumes the presence of an object. In the second stage it processes proposal and classifies the object in each area. The two-stage RPN detectors are the part of the architecture series: R-CNN, Fast R-CNN, Faster R-CNN, R-FCN, Mask R-CNN. R-CNN range detectors are very accurate, but a great problem of such networks is the low speed which upon average is only 5 FPS per GPU [12, 13].

The one-stage algorithms consist in forecasting the bounding box and the probabilities of belonging to classes for each object at once over the whole given image. With this approach one convolutional network at a time can detect many objects and their probability of belonging to the specified classes. This significantly increases the speed of operation compared to RPN. The main representatives of this group are the YOLO and SSD. In general, one-stage detectors are less accurate than two-stage detectors, but are superior in speed [8, 14, 15].

Basic architectures belong to classification convolution networks. Among the most popular architectures are: MobileNet, VGG, GoogLeNet, ResNet [16].

In order to select a detector, for the task of analyzing the customer flows distribution, studies were performed. Comparative characteristics of the studied models of the most popular pre-trained neural networks are summarized in Table 1. All models reviewed in Table 1 were trained on datasets, they include the class "Person".

Table 1
Comparative study of the pre-trained neural networks models

Model1	Scientists	mAP	FPS	Dataset
Faster R-CNN	Ren et al. [13]	73.2	5	Pascal VOC
YOLOv3	Redmon et al. [8]	57.9	20	COCO
YOLOv3-Tiny	Gong et al. [14]	33.1	220	COCO
SSD300	Liu et al. [15]	74.3	46	Pascal VOC
SSD512	Liu et al. [15]	76.8	19	Pascal VOC
MobileNet 3	Howard et al. [16]	22	31	COCO
MobileNet-SSD	Howard et al. [17]	35	56	COCO

The analysis of Table 1 shows that the R-CNN range models have high accuracy (mAP), but they have a very low video stream detection rate (FPS). YOLO in its turn provides a high FPS value, but has a lower mAP compared to other models. The MobileNet-SSD model has the most optimal parameters in FPS and mAP.

5. Analysis of Object Tracking Methods in Video Surveillance Systems

Video surveillance systems use various methods and algorithms to track the object's trajectory. According to [18] they are classified as follows:

- Dense optical flow – it helps to estimate the motion vector of all points in the video image, examples of this class of algorithms are Farneback, Horn-Schunck, and also SimpleFlow;
- Sparse optical flow – tracks the location of only a few characteristic points (representing the corners or edges of the object) on the video image, an example is the Lucas-Canada algorithm;
- Mean-Shift and CamShift determine the locations of the density function maxima;
- Kalman Filtering –allows you to get the probable positions of previously found objects in a new frame based on the history of its previous positions. On its basis the best for the date online tracker DeepSort is offered;
- Single object trackers (SOT) –assumes that the rectangle selects an object in the first frame and then tracks it in the next frame;
- Multiple object trackers (MOT) – assumes tracking multiple objects in each frame;
- Tracking algorithms built into the OpenCV library (Boosting, MIL, KCF, CSRT, MedianFlow, Mosse, Goturn, TLD).

Analysis of object tracking methods has shown [19] that at the moment the problem is particularly acute for providing continuous tracking. Most existing tracking systems do not support this functionality or they try to solve the problem by selecting an angle where the probability of overlap is minimal.

One of the main requirements for our task is the ability to track many objects of a video sequence. Also, when choosing a method, you must try to ensure maximum performance and reliability. In this paper trackers of the OpenCV library were subjected to experimental research.

6. Customer Traffic Distribution Analysis Based on Video Information

The scheme for obtaining information on the movement of customers based on vid-eo analysis is shown in Fig. 3.

At the first stage the detection algorithm (detector) works. Detection is made on the key frames of the video stream *F1, F1+step, F1+2*step, … .* In order to check whether new objects have appeared in the frame as well as to see if objects that were lost during the "tracking" stage have appeared, i.e. to correct the tracker operation. The system creates or updates a tracker with new bounding box coordinates for each detected object.

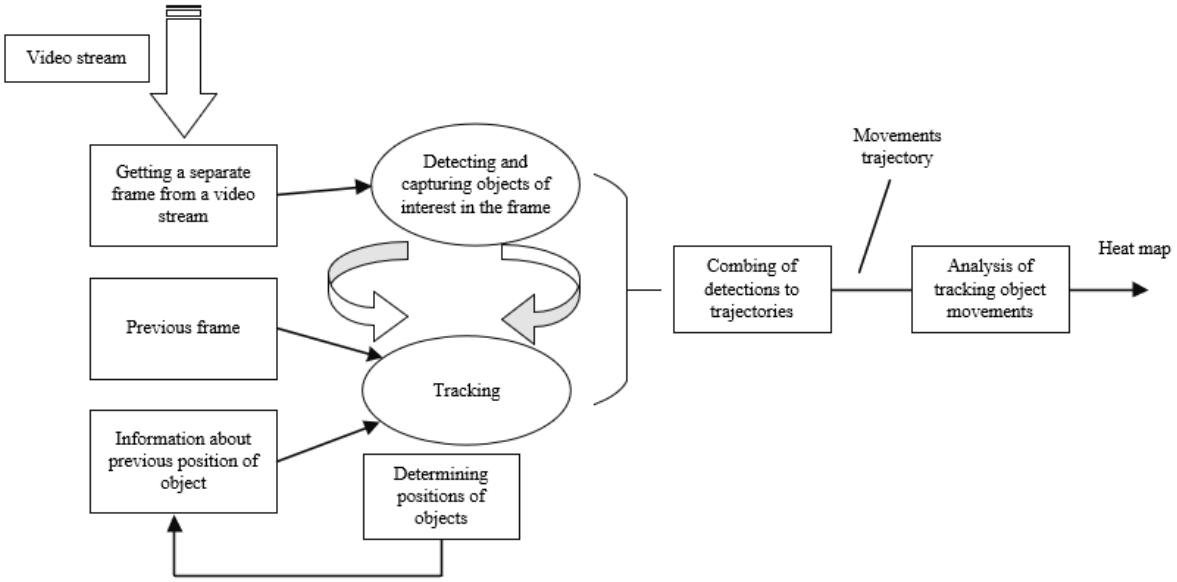

Figure 3: Generalized workflow for generating information on the movement of customers.

Since the operation of the detector allows to achieve high quality detections, and the detection algorithms are rather laborious and resource intensive, the detection phase is launched only once every *N* frames.

Each detected object is assigned a unique identifier. In the first frame identifier randomly generated and then tracked in next frames. The number of generated identifiers in the first frame is equal to the number of detections. For detections in next frames an identifier is assigned either from the previous frame (for the existing object in the system) or a new identifier is generated (for a new object).

When a new person appears in the frame, an object is created and an identifier is generated. For each of the detected objects the system creates an object tracker that tracks the object as it moves in the frame. Tracker works faster and more efficiently than the object detector. The system continues to track the object of interest until it reaches the *N* frame and then re-initializes the detector.

This problem can be considered as the construction of a matrix of scores (energies) for linking the current set of trajectories with new detections.

Trajectories linking process includes two bounding box lists: a tracking list (*t-1*) and a detection list (*t*). You need to look at the list of traces and detections with calculated IOUs (intersection over union) – a function that calculates the ratio of the area of intersection of the rectangles to the area of their union, and record the results in the matrix of scores.

According to the formed matrix, a correspondence is established between detection and tracking.

The third step is to combine detections in the trajectory. In other words a so-called optimal binding search is required: each detection joins the current trajectory or gives rise to a new trajectory. If neither the tracker nor the detector shows any bounding boxes the object is considered to be lost.

As a result of the objects identification (customers) and analysis of their movement trajectories, the time spent by each visitor in the sells areas (customer area) is determined. To facilitate the analysis of the total time spent, it was proposed to normalize the obtained time indicators:

$$V_i' = \frac{V_i - \min(V_i)}{\max(V_i) - \min(V_i)}, \quad (5)$$

where V_i – is the residence time of the object of interest in sells area *i*.

The result of this work is the construction of the heat map reflecting the areas of interest of customers, the form of which is shown in Fig. 4.

In Fig. 4 the cell at the intersection of date and department the value V_i' is indicated. The highest value of the coefficient V_i' corresponds to greater attendance by customers.

Distribution of customer flow by department
for the period from 06/15/2018 - 06/23/2018

Products \ Date	15 june	16 june	17 june	18 june	19 june	20 june	21 june	22 june	23 june
Alcoholic drink	0,1	0,1	0,2	0,1	0,2	0,3	0,1	0,2	0,3
Groceries	0,9	1	1	0,9	1	0,9	1	0,9	1
Soft drinks	0,9	0,9	0,8	0,9	0,8	0,9	0,9	0,8	0,8
Sausage and meat products	1	1	1	1	1	1	1	0,9	0,9
Confectionery	0,5	0,6	0,7	0,5	0,7	0,6	0,5	0,7	0,6
Dairy products	1	1	1	0,9	0,9	1	1	0,9	1
Cookery	0,4	0,6	0,5	0,7	0,5	0,6	0,7	0,4	0,5
Perfume and cosmetic products	1	1	1	0,7	0,8	0,9	1	1	1
Fruits and vegetables	1	1	1	1	0,9	0,9	1	1	0,9
Products for diabetics	0,4	0,5	0,4	0,5	0,6	0,4	0,6	0,5	0,4
Products for animals	1	1	0,9	0,9	1	1	0,9	0,9	1
Fish products	0,5	0,6	0,4	0,6	0,7	0,5	0,7	0,8	0,5
Snacks	1	1	1	0,8	0,1	1	1	1	1
Household chemicals	0,5	0,6	0,7	0,4	0,5	0,4	0,3	0,7	0,5
Bakegoods	0,1	0,1	0,2	0,1	0,2	0,3	0,1	0,2	0,3

hot zones — cold zones

Figure 4: Analysis of the distribution of customer flow by department.

7. Experiments

There was developed a software implementation of the subsystem in the Python programming language using OpenCV computer vision libraries and Caffe deep learning.

OpenCV [20] is an open source library for computer vision and machine learning. It contains about 2500 algorithms; the main aim of this library is to increase the computational activity of video image processing procedures.

Caffe [21] – a framework that supports the operation of convolutional neural networks. The advantage of using this tool is that it is possible to work in a very simple way and that there are pre-trained models.

Experimental research of the developed video analysis subsystem were carried out on the following computer system configuration: Windows 10 Pro, Intel (R) Celeron CPU N3050 @ 1.60 GHz, 2 GB.

The first step is choosing a detector. Models in the Table 1 are used to detect objects in the image, but they also applicable to the video stream. Since the objects detection in the video stream is reduced to processing a sequence of key frames.

The results of the experiments using standard datasets are given in Table 2.

Table 2

The results of the experiments using standard datasets

	mAP		
Model	All classes	Person	FPS
Faster R-CNN	69	78	10
YOLOv3	76	78	15
YOLOv3-Tiny	70	67	56
SSD300	80	79	35
SSD512	74	87	45
MobileNet 3	80	70	55
MobileNet-SSD	89	95	56

Analysis of Table 2 showed that the MobileNet-SSD model has the best value (mAP) for the class "Person" and a high rate of speed (FPS). The speed of the detector is important for the tracker. The lower speed of the detector, enhance the risk losing an object of interest during tracking. Thus, the MobileNet-SSD model was selected for the video analysis subsystem.

The choice of tracker was carried out experimentally using 4 test sets. Test datasets are complete video sequences obtained from shopping center surveillance cameras. Datasets contains negative parameters: various lighting, overlapping objects with each other, complete disappearance of the

object for some time, changing the size of the object. The presence of such parameters allow you to evaluate the stability of the algorithm to various emergency situations. The characteristics of test datasets are shown in Table 3.

Table 3
The characteristics of test datasets

Characteristics	Test dataset1	Test dataset2	Test dataset3	Test dataset 4
Quantity of frames	2782	654	900	525
Camera movement	No	Yes	Yes	No
Lighting	Good	Bad	Good	Bad
Overlapping	Yes	No	Yes	Yes
Complete disappearance	Yes	Yes	No	No
Changing the size of the object	Yes	Yes	Yes	No

Video analysis of the shopping center involves processing a large amount of video data, so the speed of the tracker is also important. Table 4 compares the speed of 8 trackers on four video sequences in regards to of the quantity of frames per second. According to experimental research the fastest tracker is MOSSE with an average speed of 388.38 FPS. Another effective tracker is KCF with an average value of 34 FPS. TLD and MIL trackers showed the worst results.

Table 4
Comparative study of the tracker speed

Tracker	Test dataset1	Test dataset2	Test dataset3	Test dataset 4
GOTURN	15	16	12	13
CSRT	13	23	7	34
KCF	33	41	29	33
BOOST	7	12	7	13
TLD	0.5	1.8	0.4	0.2
MOSSE	220.5	823	320	190
MedianFlow	12	35	11	45
MIL	17	12	22	19

The values of the MOTA coefficients for eight trackers on 4 video sequences are presented in Table 5.

Table 5
The values of the MOTA coefficients

Tracker	Test dataset1	Test dataset2	Test dataset3	Test dataset 4
GOTURN	20%	55%	35%	45%
CSRT	75%	65%	74%	65%
KCF	45%	25%	69%	20%
BOOST	65%	55%	75%	55%
TLD	20%	20%	35%	35%
MOSSE	56%	35%	40%	50%
MedianFlow	40%	80%	80%	20%
MIL	65%	75%	85%	50%

According to the made research it can be concluded that the BOOST tracker is very slow (with an average value of 9.75 FPS) and often loses the object of detection. MIL and KCF trackers showed

good quality of the algorithm work speed (17.5 and 34 FPS correspondingly. The TLD tracker generates a lot of false positives which makes it unusable.

The MOUSSE tracker provides the highest speed of all the considered trackers (388.75 FPS), and the CRT tracker provides a fairly high accuracy of the tracking algorithm falling short of speed herewith. The MedianFlow tracker has performed well both with regard to speed and accuracy.

8. Results and Discussion

In this work, the task of compiling customers movement map through the store was solved. There was performed the analysis of methods for detecting and tracking buyers based on video information. A subsystem for analyzing customer flows using video surveillance has been developed and its software implementation has been completed. At the first stage the problem of detection using CNN is solved. The use of MobileNet and SSD together is well substantiated. There was selected MedianFlow tracker which showed high values of speed and accuracy. The developed set of solutions made it possible to monitor the movements of customers, to identify areas of interest with the goal of further effective personnel management and display of goods.

Referenses

[1] Connell, J., Fan Q., Gabbur, P., Haas, N., Pankanti, S., Trinh, H.: Retail Video Analytics: An Overview and Survey. Proceedings of SPIE - The International Society for Optical Engineering, vol. 8663, no. 1, pp. 86630X-86630X. (2013). DOI: 10.1117/12.2008899.

[2] Hernandez, M., Nalbach, O., Werth, D.: How Computer Vision Provides Physical Retail with a Better View on Customers. IEEE 21st Conference on Business Informatics. Moscow, Russia, vol. 1, pp. 462-471. (2019). DOI: 10.1109/CBI.2019.00060.

[3] Ma, N.L., Choy, M.: Improving Customer's Flow Through Data Analytics. Advances and Trends in Artificial Intelligence. From Theory to Practice. Springer, Cham, vol. 11606, pp. 279-286. (2019). DOI: 10.1007/978-3-030-22999-3_25.

[4] Perdikaki, O., Kesavan, S., Swaminathan, J.: Effect of Traffic on Sales and Conversion Rates of Retail Stores. Manuf. Serv. Oper. Manag, vol. 14, no. 1, pp. 145–162. (2011). DOI: 0.1287/msom.1110.0356.

[5] Shengyong, Ch., Yingkun, X., Xiaolong, Zh., Fenfen, Li.: Deep Learning for Multiple Object Tracking: A Survey. IET Computer Vision, vol. 13, pp. 61–88. (2019). DOI: 10.1016/j.neucom.2019.11.023.

[6] Zhao, Z., Zheng, P., Xu, S., Wu, X.: Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11. (2019).

[7] Martynenko, T., Privalov, M., Sekirin, A.: Evolutional Approach to Image Processing on the Example of Microsections. Biologically Inspired Cognitive Architectures (BICA) for Young Scientists, Advances in Intelligent Systems and Computing. Springer, vol. 449, pp.141-150. (2016). DOI: 10.1007/978-3-319-32554-5_19.

[8] Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement. ArXiv, vol. 1804.02767. (2018).

[9] Bernardin, K., Stiefelhagen, R.: Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. Hindawi Publishing Corporation EURASIP Journal on Image and Video Processing. (2008). DOI: 10.1155/2008/246309.

[10] Szegedy, Ch., Toshev, A., Erhan, D.: Deep neural networks for object detection. Proceedings of the 26th International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, vol. 2, pp. 2553-2561. (2013). DOI: 10.5555/2999792.2999897.

[11] LeCun, Y., Bengio, Y.: Convolutional networks for images, speech, and time-series. M. A. Arbib, editor, The Handbook of Brain Theory and Neural Networks. MIT Press. (1995).

[12] Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR '14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern RecognitionJune, pp. 580-587. (2014). DOI: 10.1109/CVPR.2014.81.

[13] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no 6. Pp. 1137-1149. (2017). DOI: 10.1109/TPAMI.2016.2577031.
[14] Gong, H., Li, H., Xu, K., Zhang, Y.: Object Detection Based on Improved YOLOv3-tiny. Chinese Automation Congress (CAC), Hangzhou, China, pp. 3240-3245. (2019). DOI: 10.1109/CAC48633.2019.8996750.
[15] Liu, W., Anguelov, D., Erhan, D., Szegedy, Ch.: SSD: Single Shot MultiBox Detector. Lecture Notes in Computer Science. Springer International Publishing, pp. 21–37. (2016). DOI: 10.1007/978-3-319-46448-0_2.
[16] Howard, A., Sandler, M., Chu, G., Chen, L.-Ch., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V.: Searching for MobileNetV3. arXiv. (2019).
[17] Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv, vol. 1704.04861. (2017).
[18] Mohammad, H., Michael, G., Jonathan, S.: Combination of Mean Shift of Colour Signature and Optical Flow for Tracking During Foreground and Background Occlusion. Image and Video Technology. Lecture Notes in Computer Science, Springer, Cham, vol. 9431, pp. 1–12. (2016). DOI: 10.1007/978-3-319-29451-3.
[19] Parekh, H., Thakore, D., Jaliya, U.: A Survey on Object Detection and Tracking Methods. International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, pp. 2970-2978. (2014).
[20] OpenCV (Open Source Computer Vision Library), https://opencv.org/, last accessed 2020/04/26.
[21] Caffe: Deep Learning Framework, https://caffe.berkeleyvision.org/, last accessed 2020/04/26.