Digital Signage Personalization through Analysis of the Visual Information about Viewers

Nikolay Turov[1], Nikolay Shilov[1,2], Nikolay Teslya[2]
[1]ITMO University, St.Petersburg, Russia
[2]SPIIRAS, St.Petersburg, Russia
turovnd@niuitmo.ru; nick@iias.spb.su; teslya@iias.spb.su

Abstract – **Digital signage has been gaining more and more popularity. However, personalization of digital signage is still challenging and it is getting even more difficult with the appearance of low regulating the ways of personal data storage and processing. Most advertisement frameworks can detect general information about viewers such as age and gender that is not enough to recognize his/her real interests. There are efforts aimed at detection of emotions and prediction interests based on the mood as well as usage of modern 3D cameras for collecting distance to the viewers and their heights. Also, the topic of detection logotypes in images is getting more popular. In this paper, we are researching possibilities of integrating possible information collection techniques for better prediction of viewers' interests. These are integrated into the concept of the system that contains human detection, analytics, advertisement, and statistic modules.**

I. Introduction

Today companies are looking for new ways to attract customers attention to their products. For this purpose they use various methods like handing out flyers, putting advertising banners on vehicles or buildings. All of the methods have advantages and disadvantages. However, most of the mentioned above advertisements are static and do not change for some time extent. People get used and do not pay any attention to it.

In recent years, companies have been focusing on a digital advertisement, which is shown via large digital displays (digital signage). Such displays are usually installed in public spaces where people can look at them. Digital signage is an exciting new technology that is revolutionizing the way businesses, managers, and faculty members interact with customers, visitors, and students. The Digital Signage Association defines digital signage as "the use of electronic displays or screens (such as LCD, LED, plasma or projection) to deliver entertainment, information and/or advertising in public or private spaces, outside of the home" [1].

Many companies use outdoor advertising systems or indoor display terminals that used to broadcast predefined contents to an unspecific crowd. Digital signage systems provide their own strategies for presenting promotions like showing specific information in a certain day or time, link it to certain events, etc. There is a disadvantage of that type of systems: they do not promote personal offers. Currently, there is a growing market of customized services on personal devices, where different advertising is presented to different customers as personal devices become smart and intelligent [2].

Digital signage, which incorporates the technology for intelligent services, can provide a variety of useful information, as well as dynamic forms of advertising based on preferences of customers, improving the value of advertising and higher immersion of consumers [3].

There are lots of advertising frameworks that use gender and age recognition technologies for broadcasting suitable advertisements to some person or target group. This information cannot provide a complete overview of the person and predict his/her interests. Advertisements can be much better targeted to customers if the personal information is available. However, such regulations as GDPR in Europe and Federal Law 152 in Russia significantly limit the possibilities of personalization, since the usage of personal information is allowed only on written agreement of the person, which is nearly impossible when speaking of digital signage. As a result, the personalization can be achieved only through a deeper analysis of the viewer characteristics, e.g. height, viewing distance, face's emotions, watching time, etc.

This paper focuses on the overview of different methods for a content-based information extraction based on image analysis and suggests a concept of the system for extracting the information in real-time. The system can show the advertisement on a digital screen, detect viewer, collect base information about him/her and his/her retractions of watched ads. Moreover, the system has a model for suggestion advertisements and the analytic module for adjusting the model.

The organization of this paper is as follows. The overview of existing intelligent frameworks and different methods of detection recognizable visual images are given in Section 2. Section 3 proposes a description of the personalized digital signage system including process diagrams. The major results and future work are discussed in the conclusion.

II. Overview Of Existing Technologies

A. Overview of existing intelligent frameworks for digital signage

In recent years, intelligent frameworks for digital signage have been widely developed in different countries. Their methods, advantages, and disadvantages are discussed below.

1) Implementation of Age and Gender Recognition System for Intelligent Digital Signage

Paper [3] was written by a team from different universities in 2017. The authors developed intelligent digital signage

systems that offer targeted advertisement and information based on analyzing users. In the paper, age and gender recognition experiments were conducted using a public database. The experiment results were analyzed through a histogram matching method by extracting Local Binary Patterns (LBP) features after normalizing facial areas in input images.

The system uses pre-trained coordinates of face images based on the Active Appearance Model for recognition. Then, the normalization process is carried out by fitting the model to new input face images and finding the coordinates. Then, a histogram of LBP feature is calculated., age and gender are recognized using the nearest neighbor method. Also, there was a system test with using FG_NET and MORPH databases, that have public status.

As a result, the authors archive a high recognition rate. The gender recognition rate was 97% on average while the age recognition rate archive 68% on average.

2) Intelligent Digital Signage System Based on Gender Identification

Paper [4] was written by R. E. Abraham and M. R. Kennedy in 2018. The developed system is able to broadcast an advertisement to the digital signage display system based on the gender of the viewer. Before broadcasting appropriate content, the system analyses the demographic features of the viewers.

Gender recognition is done based on face image identification. The different face features can give information about whether a given face image is male or female. The face features of the viewer are determined by using computer vision methods. Once the gender is determined, the corresponding favorable advertisement gets displayed using database operations to the digital signage system. Real-time video of the observer is taken using the camera, and the detection of the face is done using the Haar cascade classifier. The images in the dataset are learned and computed by extracting the features of the face.

3) Intelligent advertising framework for digital signage

Paper [5] was written by a team from Intel Corporation in 2012. They draw attention to the Intelligent Advertising Framework (IAF), which uses Anonymous Viewer Analytics (AVA) and Data Mining technologies to achieve Targeted and Interactive Advertising.

Authors demonstrate an architecture of IAF that uses the following modules to predict target advertisement for the viewer:

- AVA detects the viewer and sends this information to the Analytic Server;
- Analytic Server cleans information and stores it in the Cloud when received analytical reports from the Data Mining Module (DMM);
- DMM provides data mining algorithms and a mechanism for learning and querying advertising models;
- Content Management System (CMS) takes advertising models learned by DMM and other advertising information to create customized advertising lists that are displayed.

In the paper, the authors also propose the targeted advertising methodology that includes Audience Targeting Methods and the Targeted Advertising Process.

There are three methods in Audience Targeting Methods:

- *Seeing Based Targeting:* targeting the audience once the digital sign "sees" the audience;
- *Prediction Based Targeting:* first predicts the passers coming in the future period of time and then targets them;
- *Context-Based Targeting:* the ads only depend on the context such as date/time, device location, weather info, etc.

The Targeted Advertising Process consists of the following elements:

- *Create Viewer Events*: get video and then detect humans and demographic data (gender, age bracket);
- *Learn Advertising Models:* detect viewer event with collected context information, discover viewer patterns and train advertising model with correlated data;
- *Create Default Playlist:* transfer advertising model to the CMS and extract Ad Category list from a model;
- *Finalize and Play the Playlist.*

Authors performed six experiments using two models (seeing based targeting model and context-based targeting model) and three algorithms (Decision Tree, Association Rule, and Naive Bayes). The Decision Tree algorithm achieved the highest targeting accuracy for both types of models. But the context-based targeting model got the highest results (71.72%).

4) An Intelligent Image-based Customer Analysis Service

Paper [6] was written by a team from Chunghwa Telecommunication Laboratories in 2014. In the paper, the authors focus on customer analysis. They use the RGB-D sensor that provides the 3D depth information and 2D color image recognition to the system. The 3D information needs to detect viewers, their height and distance to them. 2D color image recognition is aimed at detection of human's face and contour region, recognition of gender, age and attention time to estimate the degree of interest.

The authors illustrated two architectures of Intelligent Image-based Customer Analysis Service:

- *The Digital Signage Architecture:* when people watch at the display, the RGB-D sensor catches 3D and 2D information about of them, then the system defines interests and shows information on the screen;
- *The Store Entrance Architecture:* the RGB-D sensor is placed at the store entrance and the system defines customer information that helps the seller to make useful offers.

Moreover, the authors developed the Image-Based Customer Analysis System that uses the next components for work:

- *RGB-D Sensor:* allows retrieving information from RGB and Depth images;
- *Human Detection Method:* analyses pixels from Depth images and detects humans;
- *Human Tracking Method:* detects three events: a person enters the camera scene, a person is in the camera scene, a person leaves the camera scene;
- *Face Detection Method:* is composed of two methods Viola-Jones face detector and skin color detection;
- *Gender / Age Classification Method:* the face features are extracted by the Local Binary Pattern Method and the gender/age classification is performed by Support Vector Machine classifier.

Authors did an experiment with the Digital Signage Architecture. Their system archived 100% detection rate of people recognition, 85% gender and 75% age recognition.

To sum up, 3D human detection can increase the accuracy of person recognition compared to other methods. The RGB-D sensor is not cheap, so it requires investments for installing it to a digital screen.

5) Designing an Intelligent Digital Signage System for Business Marketing

Paper [7] was written by C. Yıldız and V. Tecim in 2018. In the paper, the authors measure interest levels of a target group to the content provided by the digital signage platform and forwarded to displays autonomously. They describe the purpose of the digital signage and highlight the benefits of implementing digital signage in business.

The authors notice that the retail and advertising industries are becoming more pervasive, they need some systems for measuring the engagement of viewers or shoppers with newly launched campaigns. They are sure that video analysis may help in understanding the audience interests. Based on their methods it is possible to detect distance, dwell time, number of viewers, attention time, glances, gender, age group by face recognition.

The authors say that audience detection systems are more usable, flexible, and cheaper than other systems. Their benefits are real-time counting, real-time detection, no human interface. They programmed an application that measures interest levels of the target group.

6) Smart Advertising Robot with Data Analytics Using Machine Vision

Paper [8] was written by a group from Malaysian, Chinese, and Japanese universities in 2017. They developed a smart advertising robot (SMADBOT) for improving the effectiveness of advertising through targeted advertising.

SMADBOT is a robot with a digital display, which provides targeted advertisements through data analysis using data obtained through machine vision. The system uses Microsoft Cognitive Services to predict emotion, gender and age group of people. Data analytics and charting are done in Microsoft Power BI by linking real-time data in SQL cloud database of Microsoft Azure. Microsoft Power BI has been used to classify consumers into various groups to achieve effective targeted advertising and let users show their own chart style and analytics. Demographic data charting of people interacting with SMADBOT has been used to let advertisers analyze the effectiveness of their advertisement.

Robot Operating System (ROS) has been used as a framework to integrate data from different sensors to perform autonomous navigation.

Several metrics have been used to evaluate the performance of SMADBOT including age, emotion prediction errors and stopping accuracy. Emotion prediction achieved a mean accuracy of 94% and had a navigation accuracy of 5.51cm in error.

SMADBOT was deployed in a real environment to further validate and test the proposed system, where 184 face counts were collected after 3-hour autonomous navigation. Data collected were successfully classified into various customer segments for effective targeted advertising.

B. Overview of methods for detection content-based information from an image

In recent years, multiple teams have developed new algorithms for deep learning logotype detection, trained data on existing image datasets and created new ones based on their needs. Some of the algorithms and datasets together with their advantages and disadvantages are described below.

1) Deep Learning Logo Detection with Data Expansion by Synthesising Context

Paper [9] was written by a team from the School of EECS, Queen Mary University of London in 2017. They introduced a new Synthetic Logos Method that allows searching for a logotype and increases the probability of its detection.

Authors formulated a new training data generation method for learning a logotype detector. They do not require large quantities of labeled training images since their model can increase the number of trained images by context variations. It is an automatically scaled synthesized logotype dataset consisting of 463 different logotypes. Also, the authors introduce TopLogo-10 dataset with logotype of high popularity.

The paper has information about the Synthetic Logos Method that includes such image manipulation methods as searching in context logotype images, using transformations (geometric, coloring) and scaling during the search for a logotype in input images.

Authors performed experiments where FlickrLogo-32 and TopLogo-10 datasets were used. They evaluated Average Precision and mean Average Precision using Faster R-CNN algorithm for logotype detection. The results showed that the probability of detection logotype using the Synthetic Logos Method increased on FlickrLogo-32 up to 10%, on TopLogo-10 up to 40%. The performance boost is obtained without the need for additional manual annotation.

2) Scalable Deep Learning Logotype Detection

Paper [10] was written by a team from the School of EECS, Queen Mary University of London in 2018. They made an overview of scalable logotype detection problem, introduce the Scalable Logo Self-co-Learning (SL2) method, create a large

WebLogo-2M dataset with 194 logotype classes and made experiments for comparing their method with existing.

For creation the WebLogo-2M dataset they used Twitter for retrieve images based on searching query. They use auto-filters to remove noise and duplicates from retrieved 4,941,317 logotype images. After filtering the WebLogo-2M dataset was created. It contains 2,190,757 logo images split to 197 logotype classes.

The SL2 method allows for discovering potential positive logotype images from web data automatically and adjust the model for improving prediction.

In the experiments, the SL2 method was compared with Faster R-CNN, YOLOv2 and other methods regarding the quality of logotype detection where input dataset was WebLogo-2M. The mean Average Precision was the compare value. The results show that SL2 archive the best value 46,9%.

III. The Concept Of A System For Extracting Content-Based Information From Video Stream

The generic system operation is as follows. The system has an interface to connect with a web camera for detection information in real-time. When someone is in the frame the camera takes an image. After processing the image, the system makes a prediction which advertisement could be interesting for them. The BPMN process of the system operation is presented in Fig. 1.

The detailed description of the system divided by modules is described below.

1) The human detection module

The aim of the module is detecting a person in a video stream. While initialization of the module, the web camera is connected for reading input video stream. The system uses YOLO v2 [11] for real-time object detection. YOLO can use different datasets for training the model. The developers suggest using convolutional weights that are pre-trained on the Imagenet [12]. The statistic page of Imagenet [13] claims that it contains more than 14 million images that are split into 5247 categories. Imagenet has more than 952 thousands of people images that are enough to detect a person and his/her face. In addition, the module should detect visual images (such as logotypes). Imagenet does not have information about them so for this purpose, FlickrBelgaLogos [14] has been used. It consists of real-world images collected from Flickr depicting companies' logotypes in various circumstances. Along with the images, it has annotations for the task of object detection. FlickrBelgaLogos has 37 logotype classes. For example, 20 logotypes of clothes include Adidas, Adidas-text, Airness, BFGoodrich, Bouygues, Bridgestone, Bridgestone-text, Citroen-text, CocaCola, Cofidis, Ferrari, Gucci, Kia, Mercedes, Nike, Peugeot, Puma, Puma-text, Reebok, and Umbro.

When the module detects a person, it starts to follow him/her. If some additional elements like company logo, tie, bowtie, T-shirt, jacket are detected with the probability of more than 80%, they are stored in the person's data. If a person's face was detected with probability more than 90% the module makes a snapshot and sends it to Microsoft Cognitive Services for detecting gender, age and emotions.

The module calculates the time between the person enters the frame and leaves it. If the person's face was recognized the attention time is calculated (time between face detection start and end).

The process of human detection and data collection is presented in Fig. 2.

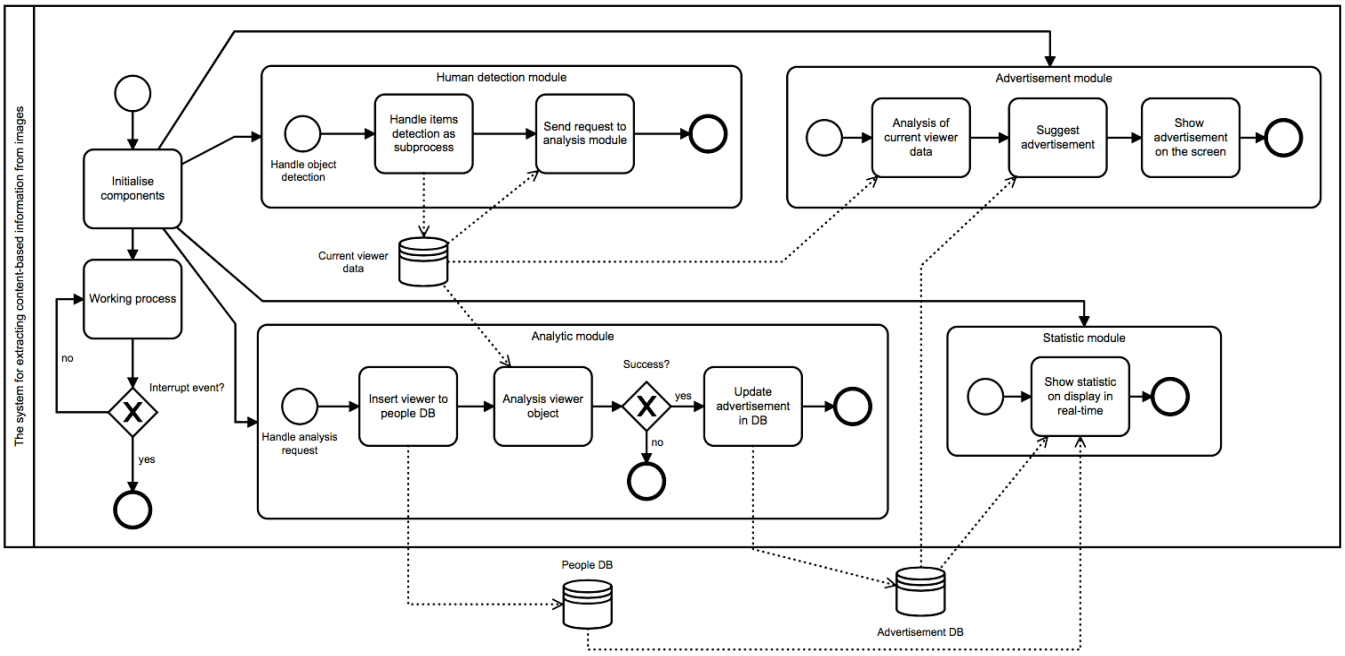

Fig. 1. The component diagram of the system

When the person leaves the video stream, the module makes a request to send the person's data to the analytics module for detecting features.

2) The advertisement module

The module is aimed at suggesting information to be presented in a digital screen (Fig. 3). The module has a model for predicting which kind of information is more interesting for the viewer. The model is adjusted when new analytic data is received.

Each advertisement information is described by the following attributes:

- *age interval*: for which age the ad is more interesting;

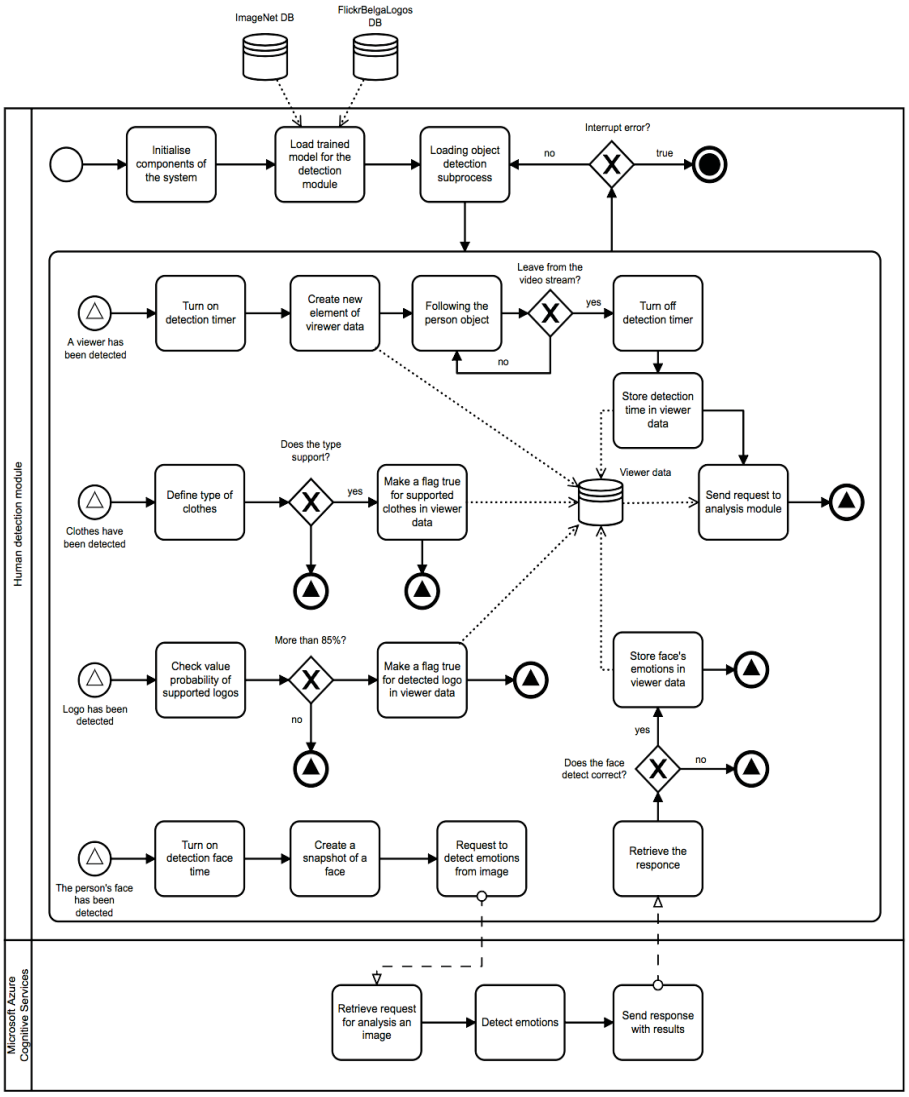

Fig. 2. The process of human detection and data collection

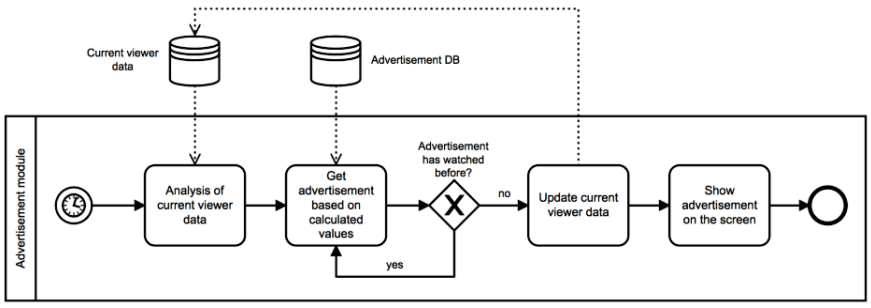

Fig. 3. The process of the advertisement module operation

- *gender:* male or female prefers the ad;
- *emotions:* what viewer feels (anger, contempt, disgust, fear, happiness, neutral, sadness, surprise) when watching the ad;
- *category:* to which category of logos (sport, business, cars, drinks) the ad matches.

The module works with the current viewer data (detected in real-time) and suggests the most suitable information for the viewer. Also, it takes into account watched information and skips it during the analysis process.

3) The analytics module

The purpose of the module is searching for features of a person's data and learning the advertisement model (Fig. 4). When the person data received for each advertisement that the person was watching the algorithm defines:

- advertisement category based on clothes and logotypes from them;
- advertisement age interval based on the person's age;
- advertisement gender based on a person's gender;
- advertisement emotions based on a person's emotions.

If an attribute is missing, it is skipped. Only defined attributes are used for updating information. The updating process consists of adding new values to specific collection's attributes and updating the score value. When the advertisement module selects an advertisement that updates take into account.

4) The statistic module

The module is designed to demonstrate statistic information in a browser. The information is split into two categories:

- Information about people include in percent:
  - gender (male or female);
  - age group (teens, youth, young adult, adult, middle-age, senior, elderly);
  - clothes (tie, bowtie, T-shirt, jacket);
  - logotypes (from FlickrBelgaLogos);

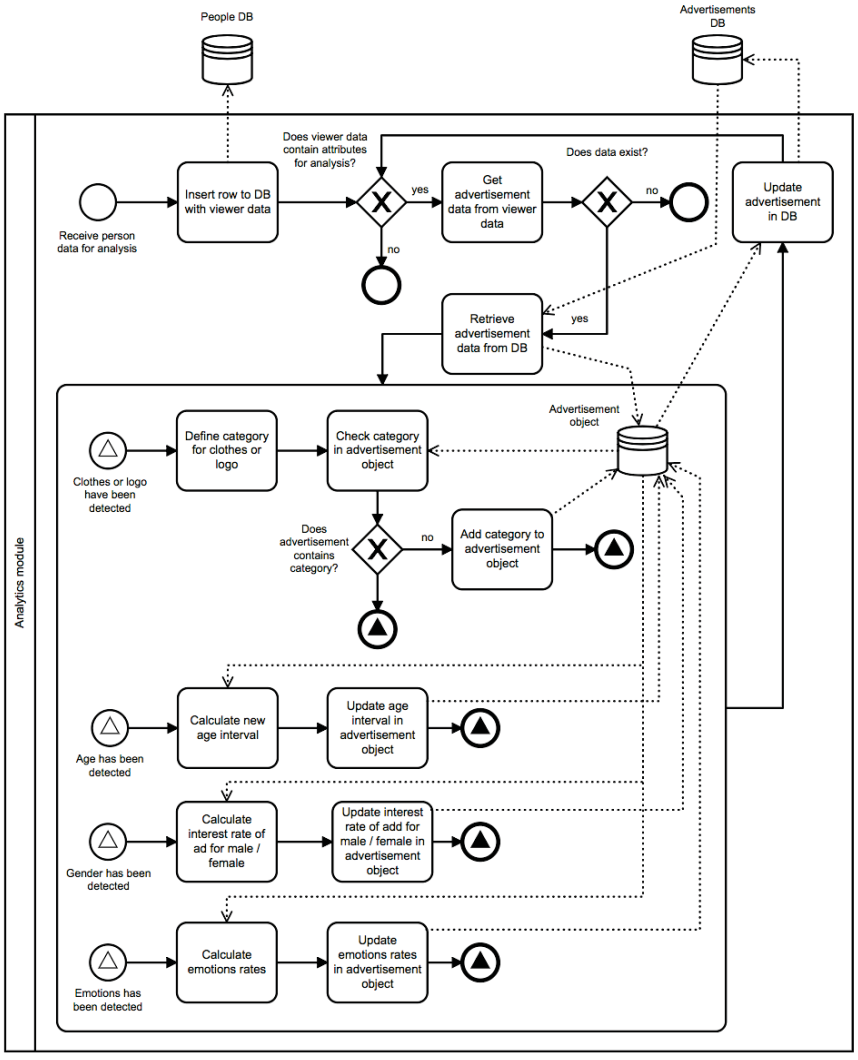

Fig. 4. The process of person data analysis

- emotions (anger, contempt, disgust, fear, happiness, neutral, sadness, surprise).

- Information about advertisement in full list with filters by:
  - gender (which ad is more suitable for female or male);
  - age group (which ad is more suitable for the selected group);
  - category (which ad is more suitable for the selected category);
  - emotions (which ad is more suitable for the selected emotion).

IV. Conclusion And Future Work

In this paper, the overview of existing advertisement frameworks and detection context-based information from the images of digital signage viewers described. The main disadvantage is that neither all of them do not try to extract content-based information like clothes or logos on them. That information can tell more about viewer preferences. The overview of methods for detection content-based information from image shows that for the last two years scientists have been looking for new ways of detection brand logotypes from images. New datasets and self-learning models for detection of logotypes have been developed.

Taking into account the analytical overview, the concept of the system for extracting content-based information from the video stream was presented. The system has 4 modules that work to predict viewer interest. In future work, the system will be developed together with underlying models, and experiments will be done.

Acknowledgment

The paper was partially supported by projects funded by grants ## 18-07-01201, 19-07-00630 of the Russian Foundation for Basic Research, by the State Research no. 0073-2019-0005, and by Government of Russian Federation, Grant 08-08.

References

[1] J. H. Meadows and A. E. Grant, Eds., *Digital signage: Communication Technology Update and Fundamentals*, 14th ed. Focal Press, 2014.

[2] S.-H. Lee, M.-K. Sohn, and H. Kim, “Implementation of age and gender recognition system for intelligent digital signage,” 2015, p. 98750I.

[3] J. Schaeffler, *Digital Signage: Software, Networks, Advertising, and Displays: A Primer for Understanding the Business*, 1st ed. Focal Press, 2008.

[4] R. E. Abraham and M. Robert Kennedy, “Intelligent Digital Signage System Based on Gender Identification,” Springer, 2018, pp. 251–262.

[5] P. Tian, A. V. Sanjay, K. Chiranjeevi, and S. M. Malik, “Intelligent advertising framework for digital signage,” in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’12*, 2012, p. 1532.

[6] T.-W. Lee, Y.-L. Chiu, Y.-S. Wu, and H.-S. Liu, “An intelligent image-based customer analysis service,” in *The 16th Asia-Pacific Network Operations and Management Symposium*, 2014, pp. 1–4.

[7] C. Yıldız and V. Tecim, “Designing an Intelligent Digital Signage System for Business Marketing,” 2018, pp. 757–767.

[8] N. A. F. M. Kamarolzaman, C. F. Yeong, E. L. M. Su, S. F. Chik, F. Duan, J. T. C. Tan, P. H. Tan, and P. J. H. Chin, “Smart Advertising Robot with Image Recognition for Data Analytics,” *Elektrika*, vol. 16, no. 3, 2017.

[9] H. Su, X. Zhu, and S. Gong, “Deep Learning Logo Detection with Data Expansion by Synthesising Context,” in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 530–539.

[10] H. Su, S. Gong, and X. Zhu, “Scalable Deep Learning Logo Detection,” 2018. [Online]. Available: https://arxiv.org/abs/1803.11417.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[12] A. Zingade, “Logo detection using YOLOv2,” 2017. [Online]. Available: https://medium.com/@akarshzingade/logo-detection-using-yolov2-8cda5a68740e.

[13] Imagenet, “Summary and Statistics,” 2010. [Online]. Available: http://www.image-net.org/about-stats.

[14] A. Joly and P. Letessier, “FlickrBelgaLogos Dataset,” 2012. [Online]. Available: http://www-sop.inria.fr/members/Alexis.Joly/BelgaLogos/FlickrBelgaLogos.html.