❒   787

AI-Based Targeted Advertising System

Tew Jia Yu, Chin Poo Lee, Kian Ming Lim, Siti Fatimah Abdul Razak
Faculty of Information Science and Technology, Multimedia University, Malaysia

Article Info	ABSTRACT

The most common technology used in targeted advertising is facial recognition and vehicle recognition. Even though there are existing systems serving for the targeting purposes, most propose limited functionalities and the system performance is normally unknown. This paper presents an intelligent targeted advertising system with multiple functionalities, namely facial recognition for gender and age, vehicle recognition, and multiple object detection. The main purpose is to improve the effectiveness of outdoor advertising through biometrics approaches and machine learning technology. Machine learning algorithms are implemented for higher recognition accuracy and hence achieved better targeted advertising effect.

Corresponding Author:

Chin Poo Lee,
Faculty of Information Science and Technology, Multimedia University,
Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia.
Email: cplee@mmu.edu.my

1. INTRODUCTION

Billboard advertising is a type of Out-Of-Home advertising that grabs the chances of outdoor promotion which typically achieves desirable results. In contrast to traditional billboards with static messages, digital billboards with more flexibility and up-to-date messages are taking their way of replacing them. This more advanced form of advertising is known as the Digital Out of Home (DOOH). As technology evolved, digital advertising becomes increasingly popular not only because of its tendency of lower cost, but also its targeting and interactive features with the use of cameras, sensors and other add-on devices.

Digital billboards are called "smart" or "intelligent" with their capabilities of recognizing a particular object and display relevant content to it. These billboards are connected to devices for collecting inputs, and a system is working behind as the control. Along with all the algorithms and processing functions, the billboard will be targeting a certain group and display relevant advertisement for better attention and influences. Several issues are identified in the existing advertising systems as the following: (1) unsuitable advertisements are displayed to the audience, (2) unable to target outdoor audience without any activity, and (3) limited functionalities. These problems result in wastage of resources, high costs in advertising, and ineffective advertising.

In view of this, an intelligent targeted advertising system is proposed for the following objectives: (1) display better-targeted advertisements to the audience, (2) improve the effectiveness of outdoor advertising, and (3) offer wide range of functionalities in a single system. The proposed system is capable of recognising gender and age of detected faces as well as various object categories such as vehicles, electronic devices, and food. Best-suited advertisement relevant to the real-time demographics will be retrieved and displayed on the billboard, solving the problems encountered in existing advertising systems as mentioned before.

The rest of the paper is organized as follows. In Section I, an introduction of billboard advertising and targeted advertising is provided. Subsequently, some existing works in smart billboard advertising are

reviewed in Section II. The system flow and functionalities are detailed in Section III. Section IV presents the experimental results of beta testing and Section V concludes the paper.

2.   LITERATURE REVIEW

Existing advertising systems are studied and a comparison is made in terms of their functionalities. Yahoo smart billboard [1] relies on a concept called grouplization, which priors on the majority to gain more attention from people around it. Yahoo smart billboard is using image recognition technology working with cameras to collect data for identification of demographic characteristics. Rather than just obtaining images, the system of Yahoo is capable of sound capturing through the use of microphones, to collect keywords spoken by a group. An additional method used to ensure attention is the eye tracking technique, detecting vision of passers-by using sensors equipped on the billboard.

NEC Digital Billboard [2] is designed specifically to display advertisements that reflect passers-by personal interests. NEC Digital Billboard uses wireless technology tags, which are also known as Radio Frequency Identification (RFID) chips. As nowadays RFID chips are increasingly being incorporated such as credit cards and mobile phones, these chips are acting like invisible labels carried by people all the way they go. The method is that these chips are encoded with information about individuals, so the digital advertising board could identify a person when they pass by, by reading the target's RFID data. NEC Digital Billboard also implemented facial recognition to identify shopper's gender, ethnicity, and approximate age.

The face-recognition billboard in London [3] is used by a global children's charity, Plan UK in their "Because I'm a Girl" campaign, to raise awareness for equal opportunity and access to education for both sexes, as well as raising fund to sponsor education for girls in developing country. The main purpose of this face-recognition billboard is to detect gender and show its entire content only to women. To achieve this, the billboard is equipped with a "high definition" camera to scan people faces, detecting their gender using face recognition technique, with a high success rate. The eye tracking technique is also used to ensure that the targeted person is looking at the billboard.

Astra Girl Detection Billboard [4] is located outside of a pub in Hamburg, Germany as part of a new advertising campaign for Astra. Rather than just focuses on promoting beer to women, the billboard even smartly avoids the youngsters under the legal drinking age of sixteen. With a built-in camera and the implemented gender-detection software, the billboard of Astra is capable of detecting the gender of people looking at it, no matter it is an individual or a group.

Lexus is moving its way to a better approach of advertising by introducing smart digital billboards to promote the cars of Lexus by triggering a personalized message to drivers corresponds to the brand, model, and colour of the vehicles [5]. In order to capture all the passing traffic, Lexus billboards rely on a series of high rotation cameras. The captured images are sent to the APN Outdoor Classifier, which is in charge of matching them to its database of vehicle makes, models, and colours, as well as the other variables. Personalized message is displayed for the targeted vehicle being recognised.

Cisco is placing this connected billboard for the intention to highlight the concept of "the Internet of everything" in advertising and to showcase its latest technology. The Cisco billboard system [6] uses a series of APIs connected to real-time traffic sensors to get the traffic conditions, in conjunction with the usage of maps and back-end network. A message with different length is displayed based on the vehicles speed, selected according to the speed range it falls into.

Together with the smart data storage company Cloudian, a Japanese advertising company Dentsu is a program corresponding to intelligent billboards has been launched [7]. The system includes the ability to analyse traffic volumes to enable highly effective targeted roadside advertising. The billboard systems are implemented using big data and deep learning. Deep leaning analysis is carried out at its first stage to attribute the recognition with automatic feature extraction of traffic patterns and volume, and also automatic vehicle recognition.

3.   INTELLIGENT TARGETED ADVERTISING SYSTEM

This section describes the system flow and training architecture of intelligent targeted advertising system.

3.1.   System Architecture

The system flow can be basically divided into four stages, namely image acquisition, multiple object detection, majority group bidding and targeted advertising. The system flow is illustrated in Figure 1.

At the initial stage, images are captured by video camera and served as input for the system. These images undergo processing in the system and are passed to the algorithms in later stage. At the second stage,

multiple object detection is deployed to detect different types of objects in a single image or frame. Detected faces and vehicles are further recognised the gender and age, or the categories respectively. The recognition processes are carried out based on the pre-trained models. An analysis is done to determine the majority group at that time point and the results are input into the bidding algorithm for advertisement selection.

The bidding output is transmitted to the database containing the advertisements in the next stage. The chosen content is retrieved from database and then transferred to the digital billboard to be displayed. The final stage is the advertisement relevant to the real-time demographic is displayed. The process iterates when the profile of detected crowd or traffic changes.
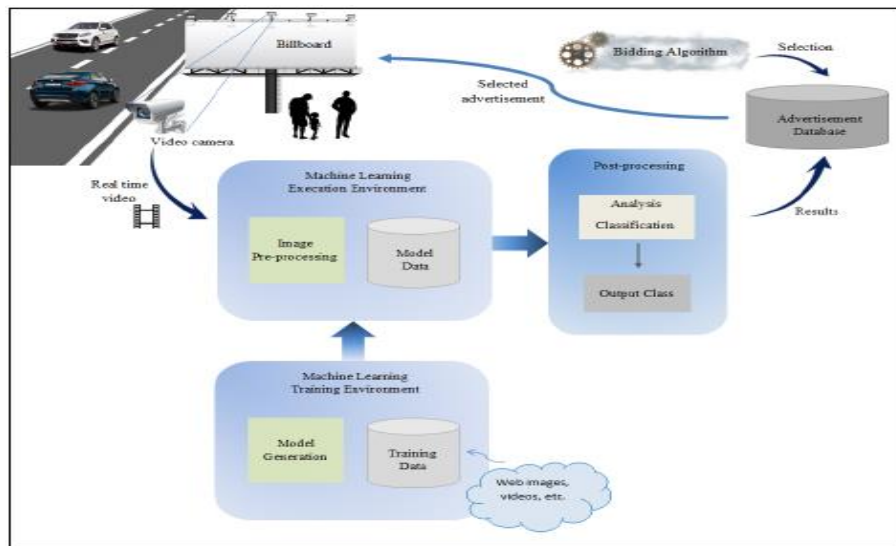

Figure 1. The system flow of intelligent targeted advertising system

3.2. Training Architectures

The models used for recognition are pre-trained before being implemented into the system. The model for recognising gender and age and the model for recognising various types of objects are trained separately by different architectures. For gender and age recognition, face cascades introduced by Microsoft is used as the face detection framework. Its function is to detect faces from camera images for gender and age recognition. The idea [8] behind this architecture is to combine face alignment with detection. Preliminary studies showed that aligned faces are able to provide better features to enhance face classification process. In the cascade framework, boosted cascade structure and simple features principles are implemented to enhance the detection efficiency. As in [9], boosting is performed on those simple classifiers, or in other words, the weak classifiers extracted are combined for better performance compared to the simple classifiers alone. In [8], the cascade detector not only takes shorter time for face detection, it also outperforms other similar solutions in detection under challenging conditions such as poor lightings, large viewpoints, and occlusion.

For the multiple object recognition, the MobileNet architecture is used to train the implemented models. As described in [10], MobileNet is a light weight deep neural networks architecture, which is built using depth-wise separable convolutions, or known as factorised convolutions. MobileNet consists of 28 layers, comprising depth-wise and pointwise convolution layers. Only the first layer of the MobileNet structure is built on full convolution just like other typically seen neural networks. In this architecture, a standard convolution is factorised into two different convolutions, namely the depth-wise convolution and the 1×1 pointwise convolution. In depth-wise convolution, input channels are filtered but not instantly combined to create new features. It requires an additional layer, which is the pointwise convolution layer to compute a linear combination of the output of depth-wise convolution via the 1×1 convolution. By having two separate layers, the computations, model size, and computational cost are much reduced.

MobileNet is used to train on Common Objects in Context (COCO) dataset for object detection and recognition in the proposed system. The COCO dataset is presented by Microsoft mainly for object recognition. The dataset contains sample photos of 91 object categories, including all the categories from PASCAL VOC ad super categories as in [11]. In this dataset, shape mask is used to detect objects using bounding box approach, providing a more accurate measure of the articulated objects.

Other than detecting and recognising objects, similarity scores of the recognised objects are computed in the system. The scoring function is implemented by using the Single Shot MultiBox Detector (SSD) architecture. In a recent study in [12], for each detected object in a bounding box, predicting scores are computed for each object category. Adjustments are then performed on the bounding box to better match the object shape. SSD is also able to encapsulate all computation into a single network. This makes SSD easy to train and less complex to be integrated into the system for detection purposes. During the SSD training, for each object involved, it only requires an input image and ground truth boxes for the detection. At each location, there is an evaluation of the default boxes of different aspect ratios with different scales. The evaluation is being processed in several feature maps [13]. These default boxes are then matched to the ground truth [14] boxes in the training phase.

3.3. Gender and Age Recognition

In gender and age recognition process, face images are acquired by the system as input data. The images are pre-processed and passed to the face detection functions. The exact face position will be computed and cropped out from the unnecessary background to optimise the recognition process. Cropped face image then undergoes the feature extraction process. Significant feature points are extracted by Microsoft face cascades algorithm to form a complete face map. The obtained face map is analysed and the output results contributed to the classification process. The system finally generates the predicted gender and age of that particular face.

In the system, gender and age recognition were implemented using Microsoft Face API. The Microsoft Face API offers a wide range of functionalities included face identification, similar face search, and face grouping. Only two face attributes are configured to return their values as required, and the return of face ID and face landmarks values are disabled. The captured images are transmitted over the Internet to the Microsoft Cognitive server for recognition, and the results are retrieved in a parsed list, which includes the parameters of gender and age. Sample results of gender and age recognition are displayed in Figure 2.

Figure 2. The sample output of gender and age recognition

3.4. Multiple Object Recognition

Multiple object recognition aims to detect and recognise different kinds of objects in real time by using Tensorflow readily trained object models. Video images are captured and passed to the system as input data. The captured images will undergo pre-processing such as resizing. Using the pre-trained models loaded into the system, significant features are extracted from the test video frames and matched with the trained object models. The extracted features are then classified as the most similar object class. Classes involved in the system are car, bicycle, motorcycle, bus, truck, animal, bag, umbrella, tie, suitcase, bottle, fruit, food, laptop, cell phone, and book.

The model implemented in the system is the SSD MobileNet COCO model. This model is trained using the convolutional neural networks on the Microsoft COCO dataset. The number of object classes defined in this trained model is 90. For detecting objects, SSD is implemented as the algorithm that detects objects in images using a single deep neural network, by putting bounding boxes over detected object features according to the feature map. The network will then generate scores for each object category in each box and further produce adjustments to the box to better match the object shape, as described in [4]. Labels for each object class is loaded into the system as a file type accessible by the Tensorflow technology.

In order to perform recognition process, the captured and saved video frame is converted into data array. A session is created for a new graph of execution and resources allocation. Necessary variables are initialised in the session. For Tensorflow computation purposes, the array of image is expanded by adding the missing dimensions required for serving the tensor as input to the functions. During the session running time, the confidence value of the detected object class is returned in a vector, corresponding to the index of class labels in the model setup process. The sample output of multiple object recognition is illustrated in Figure 3.

Figure 3. The sample out of multiple object recognition

3.5. Advertisement Selection
The selection of advertising video is based on the maximum number of the recognised object categories. As the gender and age recognition and multiple object recognition are implemented with different APIs and models, the recognition results are retrieved separately. For instance, if most adult women are recognised at a time point, advertisement relevant to this group will be displayed on the billboard. On the other hand, if the number of recognised objects is greater than the number of recognised person, the object category with the greatest number will be referred to select an advertisement related to it.

4.   RESULTS AND ANALYSIS
The models for recognition which are implemented in the system is trained using the concept of machine learning. The first step for machine learning is data acquisition. Raw data are collected and classified into three sets, namely training set, validation set, and testing set, typically with the percentage of 70, 20, and 10. All three data sets are generated randomly and consist of samples from all the output classes to ensure efficient training. The training set is used to train the models for recognition, validation set to tune the model parameters to minimise the output error rates, and testing set to assess the performance of the final model. The refined and completed models are finally placed into application. Recognition is now based on the new data from the real world.

Testing ensures the level of performance, stability, and acceptance, thus brings significant improvements and refinements to the system. Beta testing was conducted for the system by real software users to ensure that the system can handle the required and significant tasks in real-world scenarios. The system has passed all the tests and the test fields are listed in Table 1.

The test conducted on the system is divided into ten fields. As the core aspects of the system, detection and recognition functions are ranked as top priorities to be tested. There should be successful detection in an acceptable distance and the recognition accuracy for the recognition processes should be above 80 percent. For camera image capturing, video streaming, suitable advertisement displays, and proper result text displays, clear resolutions are required as well as error-free processes. The process of detection and recognition is also required to be completed in an acceptable time period.

Table 1. Results of Beta Testing

ID	Test Field	Expected Results	Pass / Fail?
1	Face detection	• Able to detect human faces in an acceptable distance	Pass
2	Gender and age recognition	• Able to perform gender and age recognitoin based on detected faces without any error • Recognition accuracy is above 80%	Pass
3	Vehicle type recognition	• Able to perform vehicle type recognition based on detected vehicles without any error • Recognition accuracy is above 80%	Pass
4	Various object category recognition	• Able to perform recognition based on detected objects without any error • Recognition accuracy is above 80%	Pass
5	Multiple object detection	• Able to detect and recognise multiple kinds of objects on the same image • Smooth process without error	Pass
6	Camera image capturing	• Captured images are clear • Smooth process without any delay or error	Pass
7	Video streaming	• Videos displayed in acceptable resolution • Smooth process without any delay or error	Pass
8	Display of appropriate advertisement	• Relevant advertisement of the largest detected object category is selected and displayed • Selection based on real-time demographics	Pass
9	Display of proper result text	• Text displayed on system interface is based on recognition results • Shows an accurate number of detected person or objects • Up to date with the real-time recognition results	Pass
10	Recognition speed	• The process of detection and recognition is in an acceptable time period	Pass

5. CONCLUSION

This paper presents an intelligent targeted advertising system that aims at providing a better advertising experience for both the advertiser and the audience. The intelligent targeted advertising system consists of several integrated functionalities, including gender and age recognition, vehicle type recognition, and multiple object detection. The multiple object detection technology provides the capabilities of detecting different kinds of objects on a single image, and further enabling the detection and recognition of human faces, vehicles, and various kinds of objects. Facial recognition is implemented to recognise gender and age based on facial features. Multiple object recognition technology is used for vehicle types and different categories of object recognition based on their unique characteristics. All of the models used in the system for recognition are pre-trained with the concept of machine learning for highly accurate results and better performance. With the system ability to display targeted advertisement content, it benefits the advertisers as they could significantly reduce their promotional costs due to the effectiveness of targeted advertising. As for the audience, they will be exposed to content that are more relevant to them, and be offered products and services that they might require.

REFERENCES

[1] Kim, E. (2016). "Yahoo has an idea for "smart" billboard that monitors people with a camera and microphone | Business Insider Malaysia". Retrieved July 19, 2017, from http://www.businessinsider.my/yahoo-patents-smart-billboard-with-camera-and-microphone-2016-10/?r=US&IR=T
[2] Digital Signage. (n.d.). Retrieved March 1, 2018, from http://my.nec.com/en_MY/solutions/digital_signage/index.html
[3] "Facial Recognition Billboard Only Lets Women See The Full Ad". (n.d.). Retrieved March 1, 2018, from https://www.psfk.com/2012/02/facial-recognition-billboard.html
[4] Hutchings, E. (2015). "Gender-Detection Billboard Advertises Beer Only to Women". Retrieved July 22, 2017, from https://www.psfk.com/2015/06/astra-beer-gender-detection-billboard-advertises-to-women.html
[5] Foote, B. (2017). "5 Eye-Popping Lexus Billboards – Clublexus". Retrieved July 22, 2017, from https://www.clublexus.com/how-tos/slideshows/5-eye-popping-lexus-billboards-468623#1-this-is-the-new-lexus
[6] Fineberg, S. (2014). "Cisco Launches Connected Billboard in San Francisco | BtoB – AdAge". Retrieved July 22, 2017, from http://adage.com/article/btob/cisco-launches-connected-billboard-san-francisco/293168/

[7]   Mellor, C. (2016). "Cloudian clobbers car drivers with targeted ads • The Register". Retrieved August 24, 2017, from https://www.theregister.co.uk/2016/06/21/cloudian_could_clobber_car_drives_with_targeted_ads/

[8]   Chen, D., Ren, S., Wei, Y., Cao, X., & Sun, J. (2014). "Joint cascade face detection and alignment". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8694 LNCS*(PART 6), 109–122. https://doi.org/10.1007/978-3-319-10599-4_8.

[9]   Ferreira, a J., & Figueiredo, M. a T. (2012). "Boosting Algorithms: A Review of Methods, Theory, and Applications". Ensemble Machine Learning: Methods and Applications, 35. https://doi.org/10.1007/9781441993267_2

[10]   Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … Adam, H. (2017). "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". https://doi.org/arXiv:1704.04861.

[11]   Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., … Zitnick, C. L. (2014). "Microsoft COCO: Common objects in context". *Lecture Notes in Computer Science (Including Subseries Lecture Notes* in Artificial Intelligence and Lecture Notes in Bioinformatics), 8693 LNCS(PART 5), 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

[12]   Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). "SSD: Single shot multibox detector". *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9905 LNCS*, 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.

[13]   Ren, S., He, K., Girshick, R., Zhang, X., & Sun, J. (2015). "Object Detection Networks on Convolutional Feature Maps", 1–8. https://doi.org/10.1109/TPAMI.2016.2601099.

[14]   Richter, S. R., Vineet, V., Roth, S., & Koltun, V. (2016). "Playing for data: Ground truth from computer games". Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9906 LNCS, 102–118. https://doi.org/10.1007/978-3-319-46475-6_7.

BIOGRAPHIES OF AUTHORS

	Tew Jia Yu was graduated from Multimedia University, Melaka holding a Bachelor of Information Technology (Honours) Security Technology. She currently lives in Muar Johor and works as an Assistant Manager in NSJ Trading. Her research interests include computer vision and security technologies.
	Chin Poo Lee is a Senior Lecturer in the Faculty of Information Science and Technology at Multimedia University, Malaysia. She completed her Masters of Science and Ph.D. in Information Technology in the area of abnormal behaviour detection and gait recognition. She is a certified Professional Technologist since 2018 and currently the Deputy Director of ADEPT (Office of Academic Development for Excellence in Programmes and Teaching), Chairperson of the Admission and Credit Transfer Committee, a senior researcher of a few Mini Fund projects, and Project Leader of an external grant project funded by MOHE. Her research interests include action recognition, computer vision, gait recognition, and deep learning.
	Kian Ming Lim received B.IT (Hons) in Information Systems Engineering, Master of Engineering Science (MEngSc) and Ph.D. (I.T.) degrees from Multimedia University. He is currently a Lecturer with the Faculty of Information Science and Technology, Multimedia University. His research interests include machine learning, deep learning, computer vision and pattern recognition.
	Siti Fatimah Abdul Razak received her B.Sc (Hons) with education where she majors in Mathematics and Information Technology and Master of Information technology majoring in Science and System Management from the National University of Malaysia in 2004. She completed her postgraduate studies in Information Technology from Multimedia University. She is currently a lecturer in Faculty of Information Science and Technology, Multimedia University. Her research interest includes rule mining, information systems development and educational technology.