

## An Integrated Computer Vision Based OOH Audience Measurement System

(Work in Progress)

Jialou Wang\*, Northumbria University, United Kingdom, [jialou.wang@northumbria.ac.uk](mailto:jialou.wang@northumbria.ac.uk)

Honglei Li, Northumbria University, United Kingdom, [honglei.li@northumbria.ac.uk](mailto:honglei.li@northumbria.ac.uk)

Shan Shan, Northumbria University, United Kingdom, [shan.shan@northumbria.ac.uk](mailto:shan.shan@northumbria.ac.uk)

### ABSTRACT

Out of Home (OOH) advertising models through deep learning method with demographical information such as gender, age, etc. While a more comprehensive model would involve fine-tuned information from audience. This paper proposed a subdivided apparel recognition model to enhance the existing audience measurement for OOH. SVM accompanied by Libra R-CNN and histogram intersection kernels is adopted alongside advertising board-mounted cameras, which obtain unprocessed data from which gender, age and other demographic features are discerned to determine viewers of particularly clothing advertising. Pervasive adoption for contactless consumer engagement, customised content display and consumer analysis is possible through the amalgamation of results, while audience measurement via digital advertising panels can be more effectively understood by OOH companies and businesses.

*Keywords:* Face recognition; gender classification; age classification; apparel classification

---

\* Corresponding author

### INTRODUCTION

Out of Home (OOH) advertising plays an important role in marketing field. A considerable proportion of businesses and companies' general spending on advertising focuses on OOH, which seeks to provide advertising material in internal (business buildings) and external (street advertising) environments to target populations (Costache, Popescu, Mocanu, & Ichim, 2020; A. Roux & Van der Waladt, 2014). Attracting possible consumers and advertising goods and services through low-cost, appealing and innovative methods is the focus of marketers, advertisers and companies due to contemporary high-tech developments. The effect of OOH advertising measurement is a crucial yet challenging activity for advertisers. The existing OOH audience measurement methods involves the computer vision method to extract precise audience measurement information from OOH advertising. Personal behaviour, demographic information as well as approximation of the present audience may all be derived through identifying, classifying and assessing individuals' faces through the computer vision techniques. Monitoring of gestures and facial expressions among viewers are the primary topics with regards to OOH advertising's audience measurement. Ethnic identity, gender, age and other demographic variables are explored in relation to facial expressions (Bekhouché *et al.*, 2015; Hjelmås & Low, 2001; Khryashchev *et al.*, 2014). Cutting edge findings relating to several competitive standards have been presented based on the identification of emotions, gender and age of viewers based on applying several deep convolutional neural networks (Dehghan *et al.*, 2017). Meanwhile, deriving a viewer's age based on their appearance in images is possible when insufficient data about their genuine biological age is available with the principal issue of developing dedicated age recognition databanks (Khryashchev *et al.*, 2014). Furthermore, audience measurement can be enhanced through the extraction of various facial features, for instance gesture, body positioning, body language, attention and in-viewing duration, as well as dwell duration (Baysal, Kurt, & Duygulu, 2010; Chen *et al.*, 2009; Ravník & Solina, 2013). The current studies for the OOH audience measurement call for a more integrated dimension with richer information, such as audience demographic dimensions, or consumers' self-identification of clothing. This research proposed and developed a subdivided apparel identification model for augmented OOH audience measurement. In this model, data related with viewers' clothing behaviour will be collected and classified according to its formal, or relaxed, or sports style and so forth. In combination with previous algorithm of a single or multiple extant gender and age algorithms, the model will provide a more comprehensive estimation of audience measurements for OOH.

### RELATED WORKS

#### Opportunity to Contact Measurement

In terms of OOH advertising efficacy's measurement, the principal index used is opportunity to contact (OTC). The sample population's trajectories are measured via satellite tracking apparatus, with highly accurate data regarding a person's position being garnered from such apparatus. Nevertheless, to provide precise findings, cautious selection of the sample is required, with significant expenditure potentially incurred through the process. Additionally, establishing the number of cell phone users walking through a location is an alternative approach, with the cell phone trajectories determined on the basis of timestamp information and associated likely position each time a cellular network is linked to by a cell phone (Quercia, Di Lorenzo, Calabrese, & Ratti, 2011). A further means of measuring the success of OOH advertising has been questionnaires, with every journey made by a sample population recorded by them. For example, the success of car advertising was investigated on the

basis of 400 questionnaires in Roux's study (2014). Moreover, the success of advertising near Hong Kong city centre metro station exits was investigated through 408 questionnaires completed by tourists and citizens in Chan and Fung's research (Chan & Fung, 2013). Regardless, precision may be markedly affected by the duration of time between the sample population's journey and when the questionnaire was completed by them.

Measuring OOH advertising success may also be done using the further significant index of eye-hit rate. Through obtaining pedestrians' location data it is possible that the size of an audience may be approximately determined via the established OTC techniques explained previously, whereas OOH advertising's direct views can be determined according to the eye-hit rate. The number of individuals looking a minimum of one time at the panel can be measured as the eye-hit rate, utilising eye-tracking technology. The correlation between hit rate and the features of an advertisement, for instance panel area, orientation and distance, is investigated through an eye-tracking technology approach devised by Barber, Sanderson and Dickenson (Barber, Sanderson, & Dickenson, 2008), with foot traffic and drivers being categorised by the researchers. One shortcoming is that marketers find information regarding group, ethnicity, gender, age and other demographics significant, which was not collected in depth during the study.

### **Audience Measurement**

Movement detection and facial detection are primarily focused on audience measurement. Furthermore, automatic suggestion of goods, clothing search and retrieval engines, clothing identification and other e-commerce deep learning methods are being implemented by fashion-associated organisations (Seo & Shin, 2019). Additionally, because numerous visual signals may be comprehended, for example group-individual proxemics, emotions, dwell duration, free gaze estimation and demographics data, critical prominence is being given to pattern recognition and computer vision technology for audience measurement. Recognition rests crucially on such dynamics as the significant internal diversity in ethnic, age, female and class groupings, as well as attained subjects limited spatial resolution.

The existing literature explores numerous face detection strategies (Feraund *et al.*, 2001; Liu & Peng, 2010; Yang, Kriegman, & Ahuja, 2002). When detecting faces in varying positions an effective approach is analysing geometrical facial features, while for discerning faces in low-quality pictures or small faces a holistic representation-based detection algorithm is superior, as a representation-focused classification of face detection techniques shows (Hsu, Abdel-Mottaleb, & Jain, 2002). Two additional subcategories of a feature-orientated strategy's formulation can be identified. Because the discernment of a face within a busy environment is a common face detection issue, colour, greyscale and such pixel features can be subject to an initial low-level assessment for visual features' segmentation (Hjelmås & Low, 2001). Within computer vision applications, the foremost primary characteristic is edges; identification of facial features is undertaken on the basis of assessing line drawings of faces extracted from images as a form of edge analysis. For instance, research offers a thorough collection of objectives for edge points' effective computation (Canny, 1986). With negligible assumptions adopted regarding the solution's form, the detector's requisite conduct may be defined through the sufficiently accurate objectives.

Additionally, features such as grey and colour data can be adopted. Facial regions adjacent to lips, pupils and eyebrows are typically lighter in shade. Low-level grey-scale thresholding enables the attainment of dark areas' extraction. Discerning faces within grey-scale photographs is possible via a neural network algorithm (Rowley, Baluja, & Kanade, 1998). Every input window can be assessed by the neural network's 'router', prior to the 'detector' network processing it. The face angle is returned by the router if a face is identified in the window. Compared with thoroughly attempting every possible orientation, such a method is markedly faster and returns a more limited number of erroneous detections, given that per image location there is just a single application of the detector network.

Detecting the presence of an object is more effective based on colour, despite a simple discernment of image features being possible through grey data. People's skin hue provides a dependable characteristic under uncontrolled, complicated conditions, as the self-adaptive face detection algorithm for discerning skin colour shows. This is because a tremendous volume of skin colour pixels were initially used in HS colour space to construct the histogram skin colour model, followed by the adoption of histogram back projection to introduce skin colour segmentation for the images, which following thresholding enables the acquisition of a binary picture of the skin colour region.

### **Classification**

Computer vision studies are placing greater emphasis on determining ethnicity, gender, age and other demographic variables. Assessing human faces for characteristic signs of ageing such as spots and wrinkles is pervasively undertaken using colour discernment and edge discernment methods. Telling apart babies and adolescents from elderly people was possible using the wrinkle index devised in (Kwon & da Vitoria Lobo, 1999), based on skin wrinkle analysis and cranio-facial development theory for determining age. Furthermore, research has devised a way of identifying more likely locations using a Marked Point Process (MPP) model, with a Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm adopted with a prior probability model to limit line segments' features (Batoov & Chellappa, 2012). Presently, discerning audience data is commonly undertaken on the basis of gender and age classification.

A face's 3D composition, inter-characteristic configural relationships, as well as local features for example skin texture, eyebrows and facial hair, are three data classifications for faces that are used to undertake gender recognition. Gender

classification techniques, face detection and skin area detection on the basis of colour space, are gender classification's principal concerns. Ravi and Wilson (2010) mention undertook experimental investigation of gender classification and face detection's relationship, while the use of manually aligned faces for gender classification via an Adaboost system has been assessed by Baluja and Rowley (2007), who altered facial image's rotation, translation and scaling to undertake a meticulous contrast of SVM and Adaboost via experimentation.

Nevertheless, in terms of viewer measurement, age and gender are fundamental demographic features, meaning that we can ask whether OOH with more precise estimations can be achieved when additional viewer data reflecting other features is derived from computer vision. For example, an image may include an individual wearing or holding numerous items of apparel, which can be identified through various methods as existing research shows. One instance is the adoption of varied previous information concerning fashion apparel such as aspect ratio, size and location, to develop an R-CNN framework (Hara, Jagadeesh, & Piramuthu, 2016). Meanwhile, deep learning to discern learned features and semantic features to engage in apparel retrieval is an original strategy devised by Liu, Zhou, and Tang (2016). YOLO2 was used as the basis of the Apparel object detection system developed by Feng *et al.* (2018), with a 56ms per image detection speed, 0.73 mean recall rate, as well as 0.839 mean precision being accomplished.

Through this study, identification of the attire classes through Libra R-CNN is undertaken, developing an apparel recognition model subdivision and providing an in-depth clarification of the process.

### METHODOLOGY

As a means of augmenting the OOH viewer measurement system, identification of the apparel classes via a subdivided apparel recognition model is the approach developed here. Accordingly, the apparel, gender and age classification model are explained here. The mining of significant and regular trends and their association rules through static databanks is the pervasive approach adopted for FP-tree, Apriori and other data mining apparatus. Through adopting association rules, clustering and classification algorithms, the mining of databanks comprising of continuous or growing, timely and real-time data will be concentrated on in this study. An individuals' mouth, nose, eyes and other major features are used to discern objects through feature-focused techniques. The location of such features in relation to each other is largely fixed, meaning that the geometric connections may be used for facial description. Nevertheless, to derive the pertinent facial features, the image must be pre-processed. Instant video footage will be obtained from a camera attachedé to ée OOH display board. Tracking individuals walking past will be undertaken through person detection, with the extent to which the tracked individuals are looking at the camera as they walk past being determined through OTC. Based on the aforementioned feature vector representations, training of the standard linear SVM (Cortes & Vapnik, 1995) will be undertaken, which is used to undertake classification. Discrete investigation of every descriptor will be made, or single long feature vectors will be formed through concatenating multiple descriptors. Accordingly, two stages of the research will be carried out to achieve this. The pre-processing phase will involve: elimination of neck, hair and other areas extrinsic to the face; automated or manual facial geometric alignment; brightness and contrast normalising, for example via histogram equalisation; lowering the pixel number through downsizing; pixel value normalising or rescaling, for instance to unit variance and zero mean.



Fig.1: Face landmarks trained on HELEN dataset

The second stage will implement Zhu and Ramanan (2012) reliable facial feature detector as a means of categorising viewers' apparel, age and gender features. The nose, mouth and eye corners are among 68 particular facial features. Alignment of the images is possible through a theoretical determination of the affine transformation, based on choosing the ideal coordinates per point. Nevertheless, practically speaking the alignment outcomes lack stability, due to flaws in face shape variability and point localisations. Moreover, the minimum window size and associated sliding step will be established via the Viola-Jones

algorithm, when an image comprises of numerous individuals. Per step, both horizontal and vertical window sliding will be undertaken for the selected window size, while application of a collection of N face recognition filters will be carried out. A face will be discerned in the specific window if a positive response is provided by a filter. The process will be ceased if the maximum size of the window is reached. If this is not the case, then the subsequent selected size for the window and related sliding step should be moved to.



Fig.2: An illustration of exact clothing retrieval



Fig.3: An example of clothing attribution recognition

A classification or regression problem may be defined for the age estimator; viewer classification on the basis of various age groups to approximate age is sufficient for our research. Evidently, even if human emulation is carried out, it is challenging to use facial images in order to approximate age. Health and numerous other extrinsic variables, as well as genetic factors, influence the distinctive aging process of individuals. SVM will be adopted accompanied by histogram intersection kernels. A process of GP multi-class classification will be undertaken on the basis of the one-vs-all method, typical for SVM classification systems, while training of the binary classifier with label regression is carried out per class  $m$ ; positive instance are all  $m$  images, while negative instances are any others that are left (Rodner *et al.*, 2017). Furthermore, classification depth adds intricacy and there are numerous apparel features, thus increasing the challenge of apparel classification. Essentially, discerning between alike classes may be a vague process, with difficulties in multi-class apparel classification. In this regard, object discernment will be carried out on the basis of the newly devised Libra R-CNN model (Pang *et al.*, 2019). Varied geometric priors will be applied to fashion apparel size, location and aspect ratios. Through applying R-CNN alongside the new balanced L1 Loss, diminishment of the disparity at the objective, features and sample level is possible. This can be shown in the equation:

$$L_b(x) = \begin{cases} \frac{\alpha}{b} (b|x| + 1) \ln(b|x| + 1) - \alpha|x| & \text{if } |x| < 1 \\ \gamma|x| + C & \end{cases} \quad (1)$$

Restriction of parameters  $\gamma$ ,  $\alpha$ , and  $b$  is on the basis of the following:

$$\alpha \ln(b + 1) = \gamma \quad (2)$$

On this basis, our experiment will establish the default parameters of  $\gamma$  and  $\alpha$ . Additionally, two classes will be determined for our apparel classification: one is the pattern and colour of apparel, for instance ‘floral’, ‘green’, ‘pink’ and so forth, while the second is the type of apparel, for example ‘jeans’, ‘skirt’, ‘hat’ and so forth. It must be pointed out that distinct apparel may be deemed identical, while identical apparel may be deemed as distinct, based on the image (Seo & Shin, 2019). Additionally, tights or leggings for example may appear ambivalent in terms of classification due to their comparable feature (Hara, Jagadeesh & Piramuthu 2016). Therefore, apparel’s heterarchical nature will be determined in the image classification algorithms, enabling a contrast with Libra-RCNN to be made, so that the images’ shortcomings can be resolved (Seo & Shin, 2019).

## EXPECT RESULT AND CONCLUSIONS

As a means of augmenting the OOH viewer measurement system, a subdivided apparel recognition model for discerning apparel categories has been posited in this research. Initially, we sought to determine the location of every passer-by's face that appeared irrespective of their positioning, as well as determining attention duration through discerning the viewer's head orientation, through applying facial detection methods, thus differing from established OOH research. Following this, the raw databank was analysed to extract apparel features in addition to gender, age and other demographic features. Accordingly, the ability to discern an individual's age group and gender, as well as whether they are orientated towards the advertising panel was possible, thus responding to the extant research shortcomings through providing relevant results. Moreover, the principal colour (blue, pink and so forth), apparel pattern (plaid, spotted, floral and so forth), as well as general style (jumper, skirt, coat, dress shirt and so forth) is discernible, responding to issues regarding apparel object data.

Based on the results being assessed, important implications for both theory and application will be derived from the research. A crucial aspect of audience measurement is the accurate identification of amalgamated demographic data such as age and gender, as well as discernment of particular apparel. Furthermore, enhanced comprehension of possible consumers is achieved when viewers of advertising can be precisely classified according to apparel subdivisions.

OOH organisations in particular will benefit from the applied contributions of this study, specifically the measurement of viewers' demographic data as part of advertainment marketing which enables the subdivided apparel recognition model's results to surpass the potential assumed by organisations. Essentially, immediate term advantages of utilising consumers' demographic data for amending marketing approaches will be derived by OOH advertising organisations.

## REFERENCES

- [1] Baluja, S., & Rowley, H. A. (2007). Boosting sex identification performance. *International Journal of computer vision*, 71(1), 111-119.
- [2] Barber, P., Sanderson, M., & Dickenson, A. (2008). Postar visibility research. *School of Psychology Birkbeck College, London*. Retrieved from <http://route.org.uk/wp-content/uploads/2017/07/Visibility-Research-Visibility-Hit-Rates-for-Poster-Panels.pdf> (December 1, 2020).
- [3] Batool, N., & Chellappa, R. (2012, October). Modeling and detection of wrinkles in aging human faces using marked point processes. In *European Conference on Computer Vision* (pp. 178-188). Springer, Berlin, Heidelberg.
- [4] Baysal, S., Kurt, M. C., & Duygulu, P. (2010, August). Recognizing human actions using key poses. In *2010 20th International Conference on Pattern Recognition* (pp. 1727-1730). IEEE.
- [5] Bekhouche, S. E., Ouafi, A., Benlamoudi, A., Taleb-Ahmed, A., & Hadid, A. (2015, May). Facial age estimation and gender classification using multi level local phase quantization. In *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)* (pp. 1-4). IEEE.
- [6] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679-698..
- [7] Chan, K., & Fung, M. S. (2013). Effectiveness of subway advertising in Hong Kong. *Journalism and Mass Communication*, 3(12), 757-767.
- [8] Chen, Q., Malric, F., Zhang, Y., Abid, M., Cordeiro, A., Petriu, E. M., & Georganas, N. D. (2009, July). Interacting with digital signage using hand gestures. In *International Conference Image Analysis and Recognition* (pp. 347-358). Springer, Berlin, Heidelberg.
- [9] Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
- [10] Costache, A., Popescu, D., Mocanu, S., & Ichim, L. (2020, June). Target Audience Response Analysis in Out-of-home Advertising Using Computer Vision. In *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)* (pp. 1-6). IEEE.
- [11] Dehghan, A., Ortiz, E. G., Shu, G., & Masood, S. Z. (2017). Dager: Deep age, gender and emotion recognition using convolutional neural network. *arXiv:1702.04280v2*. <https://arxiv.org/pdf/1702.04280.pdf>
- [12] Feng, Z., Luo, X., Yang, T., & Kita, K. (2018, December). An object detection system based on YOLOv2 in fashion apparel. In *2018 IEEE 4th International Conference on Computer and Communications (ICCC)* (pp. 1532-1536). IEEE..
- [13] Feraund, R., Bernier, O. J., Viallet, J. E., & Collobert, M. (2001). A fast and accurate face detector based on neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1), 42-53.
- [14] Hara, K., Jagadeesh, V., & Piramuthu, R. (2016, March). Fashion apparel detection: the role of deep convolutional neural network and pose-dependent priors. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1-9). IEEE.
- [15] Hjelmas, E., & Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, 83(3), 236-274.
- [16] Hsu, R. L., Abdel-Mottaleb, M., & Jain, A. K. (2002). Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 696-706.
- [17] Khryashchev, V., Ganin, A., Stepanova, O., & Lebedev, A. (2014, October). Age estimation from face images: challenging problem for audience measurement systems. In *Proceedings of 16th Conference of Open Innovations Association FRUCT* (pp. 31-37). IEEE.
- [18] Kwon, Y. H., & da Vitoria Lobo, N. (1999). Age classification from facial images. *Computer Vision and Image Understanding*, 74(1), 1-21..

- [19] Liu, H., Zhou, P., & Tang, Y. (2016). Customizing clothing retrieval based on semantic attributes and learned features. <https://pdfs.semanticscholar.org/48c9/1983b51a1d443c11e46bbd5d38bc8b7c0f11.pdf>.
- [20] Liu, Q., & Peng, G. Z. (2010, March). A robust skin color based face detection algorithm. In *2010 2nd International Asia Conference on Informatics in Control, Automation and Robotics (CAR 2010)* (Vol. 2, pp. 525-528). IEEE.
- [21] Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., & Lin, D. (2019). Libra r-cnn: Towards balanced learning for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 821-830).
- [22] Pokorny, P., Tennant, J. K., & Stevenson, P. E. (2013). Systems and methods for assessing exposure value of out of home advertising assets. *U.S. Patent Application No. 13/918,722*.
- [23] Quercia, D., Di Lorenzo, G., Calabrese, F., & Ratti, C. (2011). Mobile phones and outdoor advertising: measurable advertising. *IEEE Pervasive Computing*, 10(2), 28-36.
- [24] Ravi, S., & Wilson, S. (2010). Face detection with facial features and gender classification based on support vector machine. *International Journal of Imaging Science and Engineering*, 23-28.
- [25] Ravnik, R., & Solina, F. (2013). Audience measurement of digital signage: Quantitative study in real-world environment using computer vision. *Interacting with Computers*, 25(3), 218-228.
- [26] Rodner, E., Freytag, A., Bodesheim, P., Fröhlich, B., & Denzler, J. (2017). Large-scale gaussian process inference with generalized histogram intersection kernels for visual recognition tasks. *International Journal of Computer Vision*, 121(2), 253-280.
- [27] Roux, A., & Van der Walddt, D. (2014). Out-of-home advertising media: theoretical and industry perspectives. *Communitas*, 19, 95-115.
- [28] Roux, A. T., & Van der Walddt, D. I. R. (2016). Toward a model to enhance synergy of out-of-home advertising media integration strategies. *Journal of Promotion Management*, 22(3), 386-402.
- [29] Roux, T. (2014). The effectiveness of car advertising media: Perceptions of Generation-Y Consumers. *Mediterranean Journal of Social Sciences*, 5(20), 1877-1877.
- [30] Rowley, H. A., Baluja, S., & Kanade, T. (1998, June). Rotation invariant neural network-based face detection. In *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. 98CB36231)* (pp. 38-44). IEEE.
- [31] Seo, Y., & Shin, K.-s. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, 116, 328-339.
- [32] Yang, M.-H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58.
- [33] Zhu, X., & Ramanan, D. (2012, June). Face detection, pose estimation, and landmark localization in the wild. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2879-2886). IEEE.