An Elevator Advertisement Recommendation System Design based on Machine Vision and Deep Learning

Yiting Lei
School of Journalism and Communication
Shanghai International Studies University
Shanghai, China
0221148023@shisu.edu.cn

Ting Wang*
[1] *School of Journalism and Communication*
Shanghai International Studies University
Shanghai, China
[2] *Haina Cognition and Intelligence Research Center*
Yangtze Delta Region Institute of Tsinghua University
Jiaxing, China
tingwang1981@163.com

Haibo Zhang
School of Journalism and Communication
Shanghai International Studies University
Shanghai, China
zhanghaibo2023@163.com

Tongtong Ding
School of Journalism and Communication
Shanghai International Studies University
Shanghai, China
2801559255@qq.com

***Abstract*—In recent years, machine vision sensors have been continuously developed and gradually applied to advertising industry for accurate advertisement recommendation. However, because of the isolated environment, it is very difficult to recommend advertisement in elevator cages. In this study, a novel advertisement recommendation system is designed for the elevator, based on machine vision sensors, Raspberry Pi, deep learning neural networks, information management system, and socket communication, which are employed to analyze the features of the passengers in the elevators and recommend the most suitable advertisements for them. This system aims to help advertisers improve the effect of advertising activities. The simulation results show that the proposed system in this paper has a very accurate performance and can be widely applied in the elevators.**

Keywords—machine vision, deep learning, elevator advertising, precise recommendations

I. Introduction

Nowadays, advertisement videos are firstly stored in a server. They are broadcasted in the elevator cages in many cities and displayed one by one in the form of rotation. Because most of the elevator cages are isolated space with weak mobile signals, elevator advertisements can attract more attentions of passengers and have a higher advertising arrival rate than other public advertisement videos. However, the current elevator advertising is set by the advertisers, which has less personal preference and lacks of pertinency. As a result of that, the advertising effect and conversion rate of current elevator advertising is not ideal. In this paper, a novel elevator advertising system is designed, which aims to display the optimal advertisement videos for the passengers based on their personal preference, where the elevator advertising videos are recommended to passengers by an intelligent recommendation algorithm based on their genders and ages estimated by their face information captured by machine vision sensors. Apart from that, an information system is also designed as a management platform to manage the devices, algorithm rules and advertising films.

*Ting Wang is the corresponding author.

The rest of the paper will be arranged as follows: Section 2 will review the relevant research of elevator advertising system. Section 3 will introduce the system design. In section 4, the simulation experiment and results will be demonstrated. Finally, the conclusion will be drawn in the last section.

II. Literature Review

At present, elevator advertising business has been widely carried out. Such as Germany ThyssenKrupp elevator used interactive screens. Otis elevator has also used elevator projection advertising to improve passenger immersion. Although there is a certain innovation in the above elevator advertising methods, none of them can guarantee the effective arrival rate of the advertisement. For individual users, precise recommendation algorithms of PC and cellphone applications have been relatively mature. However, for the collective users, for example, passengers in the closed environment of the elevator, obviously cannot meet the demand.

Aiming at the problem that traditional methods cannot reach accurately, Li adopts the bipartite graph model as the network representation learning method, which decomposes the user and the advertisement content into two networks respectively. [1] However, when dealing with large-scale datasets of elevator advertisements, the complexity of the algorithm may lead to high computational costs in practical applications.

Shang et al. [2] proposed a video recommendation algorithm based on the hyperlink graph model. By analyzing users' transfer patterns between different web pages, this algorithm is able to infer the types of videos that users may be interested in. However, this recommendation algorithm may not be accurate in analyzing the results for a multi-person system in an elevator. In addition, Kim et al. designed and implemented an ontology-based PTA system for iTV environments. The system utilizes semantic relations in ontology to enhance advertisement recommendation capabilities and improve the efficiency of information reuse. [3]

Overall, although the above algorithmic models provide useful attempts and results in the study, the effect for the closed elevator environment is not yet known, and there is still a need for more in-depth research and improvement in several aspects.

In face recognition systems, the accuracy of gender and age estimation is crucial to enable effective personalized recommendations. For example, a young user group is more inclined to watch entertainment-related video content, while older users are more concerned with health information. [4] The process of face recognition techniques involves complex image processing and pattern recognition techniques. For example, Cui et al. [5] proposed a video recommendation algorithm that combines video content and social network information. In this area, Convolutional Neural Networks (CNNs) have shown excellent performance in processing visual data. [6]

Wang et al. proposed a new approach to ameliorate the problems of high computational effort, long operation time, and low accuracy that exist in traditional and lightweight face recognition algorithms. Further research could explore how these algorithms can be applied to elevator advertisements and how to improve the algorithms' generalization capabilities. [7]

On the mobile or Web side, Zhu et al. proposed a multilayer recommendation method based on path integration enabling the recommender system to understand the user's needs more comprehensively.[8] However, this multi-layer data fusion based recommendation approach also faces the challenges of real-time and personalization. [9]

III. System Design

In our study, the proposed system is divided into three parts: hardware devices, management platform, and recommendation system, as shown in Fig. 1.

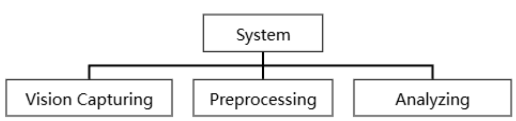

Fig. 1. Hardware function architecture.

When the system works, the client captures a real-time video stream inside the elevator by a vision sensor. Because of the limitation of computing resource, such as CPU and memory, the client cannot process all the captured visual data by itself, otherwise the operating efficiency of the system would be very low. Therefore, data should be pre-processed on the client side and transmitted to the server side by network communication. After data analysis based on deep learning and intelligent recommendation algorithm, the server put out the optimal advertisement for passengers in the elevator, and delivers the corresponding advertisement video. The management platform manages and adjusts raw data, clients, advertisements, recommendation rules, users, etc.

A. Hardware

In the elevator, the system uses a camera as a vision sensor to capture real-time facial data of passengers and convert it into digital signals for computer processing, based on Raspberry Pi 4B in this study, as shown in Fig. 2.

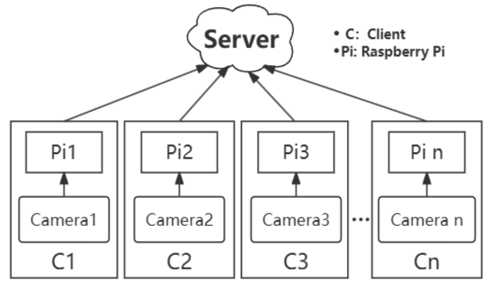

Fig. 2. System architecture.

For data transmission, the system uses a scheme where data is preprocessed by the client before being sent to the server. Specifically, the facial data captured by the vision sensor is initially processed by the Raspberry Pi and then transmitted to the server over the network by socket communication. Importantly, the system does not store any facial data locally. This design not only mitigates the risk of privacy breaches due to local storage but also reduces the likelihood of data loss or tampering. All data is securely stored on a reliable server, ensuring data integrity and security.

Socket programming enables two-way communication between different computers, making it ideal for real-time data transfer scenarios. In this communication mechanism, the client and server exchange and transfer information through a socket connection. The workflow of the client and server programs is shown in Fig. 3.

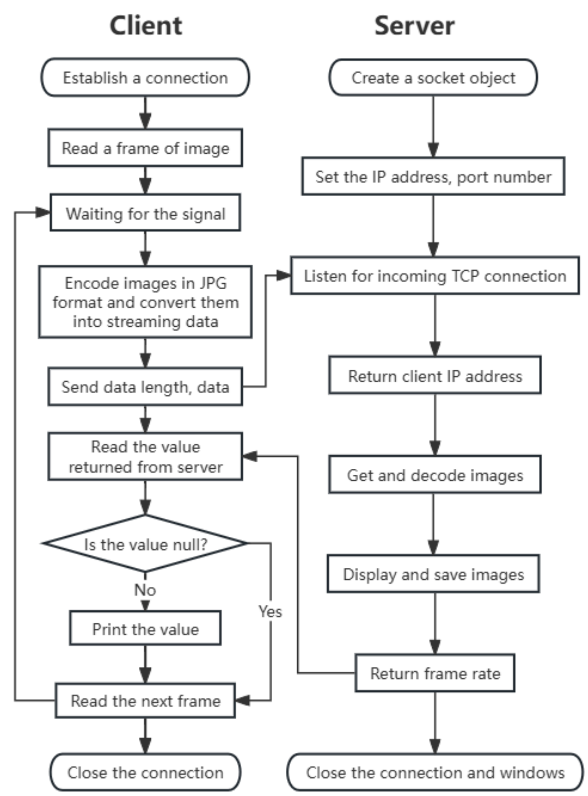

Fig. 3. Client - and server-side program flow.

After the Raspberry Pi starts, it initializes the socket connections. In the elevator's interior environment, once a passenger enters, the vision sensor immediately captures their facial data and sends it to the Raspberry Pi in real time. Upon receiving the facial data, the Raspberry Pi promptly performs initial processing and encapsulation. The processed data is then sent to the server over an established socket connection. The server, upon receiving the data from the Raspberry Pi, intercepts a passenger image every three seconds based on preset rules. This facilitates subsequent facial recognition and advertising recommendations.

B. *Management platform*

The platform is designed to provide a user-friendly interface, enabling administrators to easily monitor the system and manage advertising content, as shown in Fig. 4.

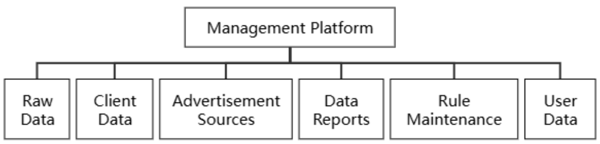

Fig. 4. Management platform functional structure.

The management platform enables real-time monitoring of the elevator advertising system's operational status, assesses the effectiveness of advertisements, generates statistical and analytical reports on playback data, and aids administrators in understanding the audience and delivery impact of advertisements. Additionally, administrators can upload new advertising materials, edit existing content, monitor advertising effectiveness, and optimize strategies based on data insights to enhance advertising impact and user satisfaction. The page layout is shown in Fig. 5.

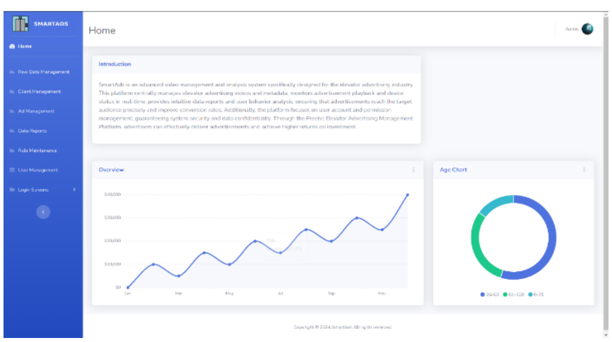

Fig. 5. Management platform home page.

C. *Recommendation system*

The entire algorithm flow of the system involves real-time face recognition and advertising recommendation, as shown in Fig. 6. A deep neural network model is employed to identify gender and age, and advertising recommendations are optimized based on the recognition results. The algorithm structure can be summarized in the following steps:

- Initialization Phase: Import the necessary libraries. Define three auxiliary functions: one for converting text for gender and age into numeric representations, another for finding the maximum value in a dictionary and returning the corresponding key and value, and the third for detecting faces in images and returning the image with the face box and the coordinates of each box.
- Model loading phase: Load pre-trained deep neural network models, including models for face detection, gender recognition, and age recognition. For face detection, the DNN (Deep Neural Network) forward propagation algorithm is utilized, while gender and age recognition employ pre-trained network models. Define the model mean. The model used in the code is the standard model, with no special modifications.
- Video processing and real-time recognition phase: Open the video file. Process each frame image as follows: Detect faces in the image. For each detected face, make predictions using a gender and age recognition network. Convert the predicted results (gender and age) into numerical representations and save them to a list. Draw a face frame and the corresponding gender and age labels on the image.
- Advertising recommendation stage: Based on the identified gender and age group, retrieve the corresponding list of advertising preferences. Sum up the liking scores for each advertisement type. Select the top three ads with the highest scores and add them to the playlist.

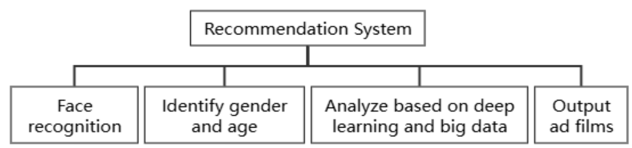

Fig. 6. Recommendation system architecture.

The algorithm structure flowchart is shown in Fig. 7.

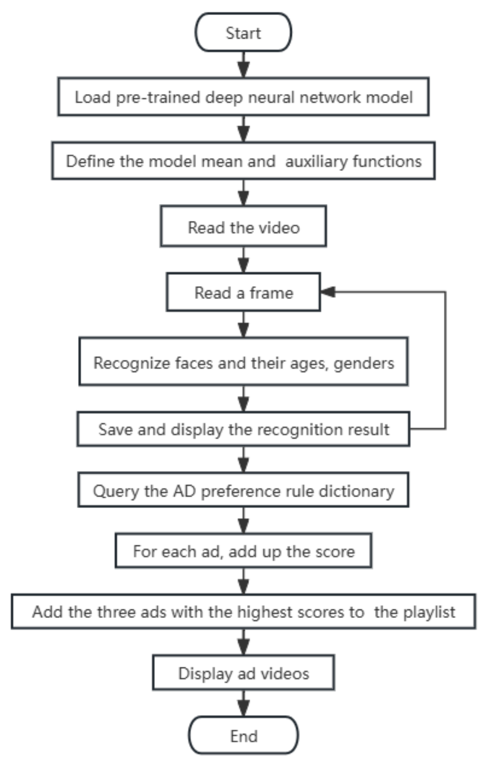

Fig. 7. Algorithm structure flow.

IV. SIMULATION EXPERIMENT

This section validates the facial identity recognition system based on machine vision principles proposed in Section 3 through experimental testing.

A. Software and Hardware Configuration

The proposed system consists of a Raspberry PI 4B, a display with an HDMI interface, a computer, and a monocular camera. The monocular camera is used to record video inside the elevator. Computer with a Windows 64-bit operating system and a 11th generation Intel(R) Core(TM) i5-11300H@3.10GHz 2.61 GHz processor, GeForce MX450 graphics card, 16GB memory, is a server. Python 3 is used as programming language and OpenCV 4.8.1 is also employed.

B. Experimental Design

The simulation experiment in this section tests the facial identity recognition system based on machine vision principles proposed in Section 3. The pixel of the test videos used in the simulation experiment is 1080*720, and the frame rate is 30. There are 6 test videos in total, which are shot by the camera in fixed positions in the elevator. The duration of each video is 30 seconds or so, and the video numbers are sequentially marked as 1, 2, 3...6. When the elevator closes, we shoot videos. When the elevator opens, we stop shooting. In the experiment, there are a total of 178 faces as facial targets.

C. Analysis of Experimental Results

To validate the effectiveness of facial identity recognition in the video, a captured image is randomly selected. As depicted in Fig. 8, three individuals looking at the camera in the video are accurately framed by green rectangular boxes. Each box displays a label in the top left corner, indicating the identified identity. For instance, from left to right, the first person's face is framed, and the label shows 'Female; 18-25', indicating identification as a woman aged 18-25. Similarly, the second person's face is framed with the label 'Female; 18-25', and the third person's face with 'Male; 18-25', indicating their respective identifications as a woman and a man, both aged 18-25. The top left corner of the screen displays 'video playing: Xiaomi-digital', indicating the digital advertisement being shown.

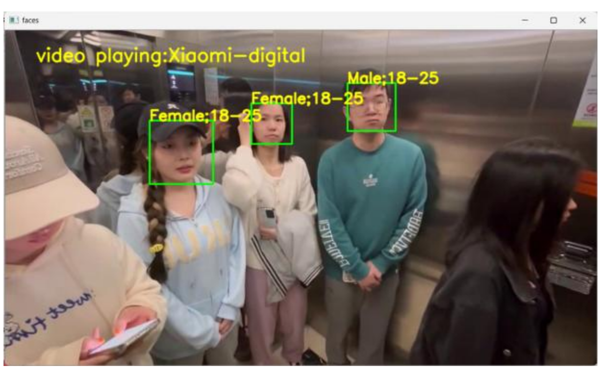

Fig. 8. Facial identity recognition capture.

The calculation formula for facial recognition accuracy is usually obtained by calculating the confusion matrix. The confusion matrix is a 2*2 matrix, where the rows represent the predicted results (Positive and Negative) and the columns represent the actual results (Positive and Negative). Specifically, the four values in the confusion matrix represent:

- True Positive (TP): the number of samples predicted as positive and actually positive.
- False Positive (FP): the number of samples predicted as positive but actually negative.
- False Negative (FN): the number of samples predicted as negative but actually positive.
- True Negative (TN): the number of samples predicted as negative and actually negative.

Through these four values in the confusion matrix, metrics such as accuracy, precision, recall, and F1 score can be calculated for pedestrian recognition. The specific calculation formulas are as Eq. (1), (2), (3) and (4):

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ samples} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$Precision = \frac{number\ of\ true\ positive\ predictions}{number\ of\ all\ positive\ predictions} = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{number\ of\ true\ positive\ predictions}{number\ of\ all\ positive\ cases\ in\ reality} = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (4)$$

Here, *Accuracy* reflects the proportion of correct predictions in all prediction results. *Precision* reflects the proportion of true positives among all positive predictions. *Recall* reflects the proportion of true positives among all actual positives, and *F1* score comprehensively evaluates precision and recall.

In this test, the average performance indicators of the whole facial recognition part of the system are shown in Table I.

TABLE I. AVERAGE EVALUATION OF THE WHOLE FACIAL RECOGNITION

Category	Accuracy	Precision	Recall	F1
Face	46.90%	85.99%	50.09%	0.63
Gender	44.27%	74.42%	51.13%	0.58
Age	18.15%	35.49%	27.04%	0.30

The gender and age recognition algorithm model for faces draws on a project published by Gil Levi and Tal Hassner[10]. The training dataset was published by Eran Eidinger, Roee Enbar, and Tal Hassner[11]. The total number of photos is 26580. The total number of subjects is 2284. The reasons for errors in facial recognition are analyzed, mainly including: 1) The low light in the elevator makes facial features indistinguishable. 2) The percentage of Asians in the training data set is not high enough, so the accuracy of identification is not high.

Performance metrics for the facial recognition component of our system in this test experiment is shown in Table II, with an average accuracy of 46.90%, precision of 85.99%, recall of 50.09%, and an F1 score of 0.63. Performance metrics for the gender recognition component of our system in this test experiment is shown in Table III, with an average accuracy of 44.27%, precision of 74.42%, recall of 51.13%, and an F1 score of 0.58. Performance metrics for the age recognition component of our system in this test experiment is shown in Table IV, with an average accuracy of 18.15%, precision of 35.49%, recall of 27.04%, and an F1 score of 0.30.

TABLE II. EVALUATION OF FACIAL RECOGNITION RESULTS

Video ID	Sample number	Accuracy	Precision	Recall	F1
1	36	63.89%	92.00%	67.65%	0.78
2	18	66.67%	92.31%	70.59%	0.80
3	12	33.33%	80.00%	36.36%	0.50
4	20	50.00%	90.91%	52.63%	0.67
5	16	37.50%	85.71%	40.00%	0.55
6	20	30.00%	75.00%	33.33%	0.46

TABLE III. EVALUATION OF GENDER RECOGNITION RESULTS

Video ID	Sample number	Accuracy	Precision	Recall	F1
1	36	47.22%	89.47%	50.00%	0.64
2	18	61.11%	91.67%	64.71%	0.76
3	12	33.33%	80.00%	36.36%	0.50
4	20	50.00%	90.91%	52.63%	0.67
5	16	43.75%	77.78%	50.00%	0.61
6	20	45.00%	90.00%	47.37%	0.62

TABLE IV. EVALUATION OF AGE RECOGNITION RESULTS

Video ID	Sample number	Accuracy	Precision	Recall	F1
1	36	19.44%	36.84%	29.17%	0.33
2	18	11.11%	16.67%	25.00%	0.20
3	12	8.33%	25.00%	11.11%	0.15
4	20	15.00%	30.00%	23.08%	0.26
5	16	25.00%	44.44%	36.36%	0.40
6	20	30.00%	60.00%	37.50%	0.46

Upon initialization, the recommendation system includes a pre-set list of three advertisements, each with a duration of 15 seconds. During the system's initial run, it first plays the advertisements from this pre-set list. Once these advertisements have been displayed, the system proceeds to play ads recommended based on the identification results.

The system begins by reading the first frame to perform facial recognition and then plays the recommended advertisement. During the playback of the advertisement, the system continues to read subsequent frames, refining and optimizing ad recommendations based on the ongoing identification results.

When the first face is recognized, record the time as $t1$. When all faces have been recognized and the advertisement is about to be played, record the time as $t2$. The response time is the difference between $t2$ and $t1$, measured in seconds. The formula for calculating the *Response time* is shown in Eq. (5):

$$t = t2 - t1 \tag{5}$$

The response time evaluation results are shown in Table V.

TABLE V. RESPONSE TIME EVALUATION RESULTS

Video ID	1	2	3	4	5	6
Video Duration (s)	155	88	33	29	8	10
Response Time(s)	0.026	0.027	0.028	0.032	0.003	0.028

V. CONCLUSION

The accurate recommendation of elevator advertising has always posed a formidable challenge. The elevator environment is unique, with high passenger mobility and diverse individual needs and interests. Achieving precise advertising in this context has become an urgent issue in the advertising industry. This study is dedicated to addressing this challenge by socket communication, vision sensor technology, deep learning, and data analysis. Based on these technologies, an efficient advertising recommendation system is constructed, providing advertisers with more precise and effective advertising strategies while enhancing advertising effectiveness and monitoring efficiency.

This study introduces novel ideas and methodologies to the advertising industry, significantly enhancing the effectiveness of advertising delivery and monitoring. The research outcomes not only offer practical value for advertisers but also pioneer new avenues for applying deep learning and big data analysis in advertising. Looking ahead, further exploration into the application of intelligent recommendation algorithms in outdoor advertising is anticipated. Additionally, the potential of emerging technologies such as artificial intelligence and machine learning in enhancing advertising delivery effectiveness and monitoring efficiency will be investigated. As technology advances and application scenarios expand, the accuracy of elevator advertising recommendations is expected to mature and improve, presenting both opportunities and challenges for the advertising industry.

REFERENCES

[1]. Chunhui, L., An Advertising Recommendation Algorithm Based on Deep Learning Fusion Model. Journal of Sensors. 2022.

[2]. Shang, S., et al., A Video Recommendation Algorithm Based on Hyperlink-Graph Model. International Journal of Software Innovation, 2017. 5(3): p. 49-63.

[3]. Kim, J. and S. Kang, An ontology-based personalized target advertisement system on interactive TV. Multimedia Tools and Applications, 2011. 64(3): p. 517-534.

[4]. Science - Computational Intelligence and Neuroscience; New Findings Reported from Sapienza University Describe Advances in Computational Intelligence and Neuroscience (Gender and Age Related Effects While Watching TV Advertisements: An EEG Study). Science Letter. 2017.

[5]. Cui, L., et al., A video recommendation algorithm based on the combination of video content and social network. Concurrency and Computation: Practice and Experience, 2016. 29(14).

[6]. Mısır, M. and M. Sebag, Alors: An algorithm recommender system. Artificial Intelligence, 2017. 244: p. 291-314.

[7]. Wang, S., A Face Recognition Method based on Lightweight Neural Network and Multi Hash Recognition Degree Weighting. IAENG International Journal of Applied Mathematics. 2024. 54(3).

[8]. Zhu, H., et al., A Cross-Curriculum Video Recommendation Algorithm Based on a Video-Associated Knowledge Map. IEEE Access, 2018. 6: p. 57562-57571.

[9]. Mostafa, D. and K. Hossein, Optimizing Deep Neural Networks for Face Recognition to Increase Training Speed and Improve Model Accuracy. Faculty of Electrical Engineering, Shahrood University of Technology ,Shahrood, Semnan, P.O. Box 3619995161 ,Iran. 2024. 38(3): p. 315-332.

[10]. Levi, G. and T. Hassncer. Age and gender classification using convolutional neural networks. in 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2015.

[11]. Eidinger, E., R. Enbar, and T. Hassner, Age and Gender Estimation of Unfiltered Faces. IEEE Transactions on Information Forensics and Security, 2014. 9(12): p. 2170-2179.