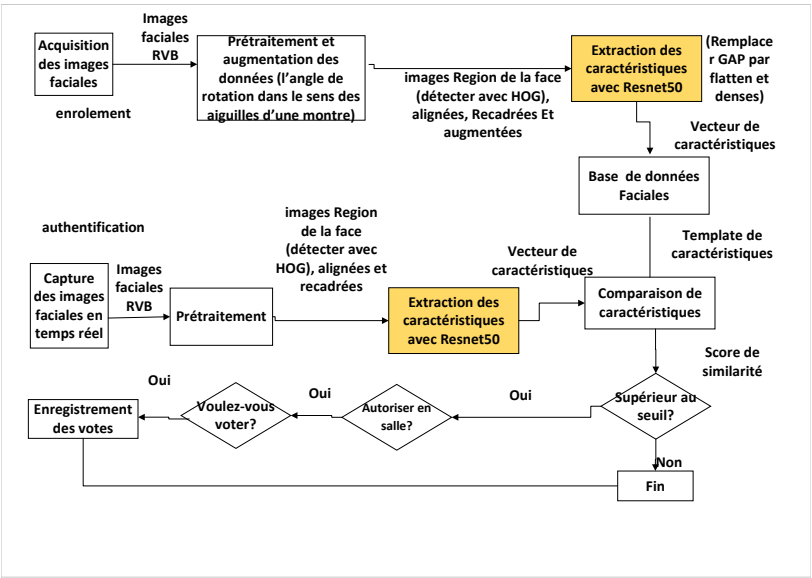

Fig. 1: A voting system based on facial recognition

This work proposes an anti-fraud system for voting.
Facial images of each voter are compared to those recorded in the database.
A positive comparison means permission to vote.

Optimization of ResNet50 based on flatten and dense layer insertion applied to facial recognition voting system

Dr. Aminou Halidou[1,2*], Tchana Ngninkeu Gil Thibault[1†] and Dr. Daramy Vandi Von Kallon[2†]

[1*]Department of Computer Science, University of Yaounde I, Yaounde, 337, Cameroun.
[2]Department of Mechanical and Industrial Engineering Technology (DMIET), University of Johannesburg, Johannesburg, 524, South Africa.

*Corresponding author(s). E-mail(s): halidou.aminou@facsciences-uy1.cm;
Contributing authors: gilthibault5@gmail.com; dkallon@uj.ac.za;
†These authors contributed equally to this work.

Abstract

Biometrics has been used in the Cameroonian electoral system for the past decades in order to improve voters' identification by delivering voter cards. During the electoral phase, voters with voter cards are identified by means of a paper-based system at their polling stations. When the results are released, opponents loudly criticize the results citing electoral fraud. One of the frauds to be resolved is multiple voting. To resolve this problem, this paper proposes a system that takes as input facial images from each voter using deep learning approach, compares it with the recorded input features, and produces results as output. Positive results imply that he/she can vote. After voting, the voter's facial image is captured again and a short text message is delivered as confirmation from the system. A summary of the votes is produced which guarantees that a person has voted only once. This paper describes the methodology adopted to build the system. The facial recognition accuracy reached 99.56% which is an improvement compared to our baseline.

Keywords: DLIB, deep learning, voting machine, facial recognition, LFW

1 INTRODUCTION

In recent years, several countries such as the United States and India has become interested in the use of biometrics. Biometric technologies such as iris, fingerprints, voice, and facial recognition offer many advantages for authentication and identification. Using these advantages, they have proposed and designed voting machines so that their citizens can vote democratically and securely. Voting is an act by which a citizen participates in the choice of his/her representatives or in decision-making during a ballot (presidential, parliamentary, or municipal elections). In view of the importance of voting, Friedrich Ebert Stiftung suggests that Africans, and Cameroonians in particular, can only contribute to nation-building and a democratic society through a free expression of political will, which is a fundamental building block to democracy [1]. In Cameroon, the electoral body is ELECAM, created in 2006 by law N° 2006/O11 of 29 December 2006. This body is equipped with fingerprinting biometric equipment (as of 2012) [2]. During the voting phase, voters go to their polling stations to exercise their voting rights (cast their votes). Each station has a paper identification system to identify the voters in detail. This system has proven its worth, but at the end of each election cycle, the opposition loudly criticizes the results of these elections (during the electoral and post-electoral phases), citing electoral fraud (filled ballot boxes, falsification of results, multiple voting by one voter) [3] [4] [5]. Here, we are interested in multiple voting by one voter. This could occur during the voters' identification simply because the system in place does not control the fact that a voter can't vote several times, at several polling stations. The persistence of electoral fraud discredits elections in Cameroon; it leads to a lack of interest from the population and hinders the development of the country, as the legitimacy of leaders to carry out their functions is contested. It can also be a cause of conflict and chaos observed in most parts of the world. It is important to organize transparent, free and fair elections.

In this paper, we propose a face recognition-based voter authentication system that captures an input facial image of each voter using deep learning approach, RESNET50 verifies it with the characteristics of the facial images that have been recorded, the result being positive means that he can vote [6]. After voting, the system confirms the vote by capturing the voter's facial image again and delivers a short text message. A summary of the votes can be consulted and thus guarantee that a person has voted only once.

2 RELATED WORKS

To mitigate electoral fraud, several works have focused on facial recognition. Manual feature extraction from facial images using traditional methods such as LBP and HOG and then use these features in the classification step [7]. Recently, the artificial neural network has been used to extract features from facial images, using large amount of dataset. Given its relevance, several

researchers continue to explore this technique. Yuxiang et al. used neural networks in particular ResNet50, which is composed of 50 layers with the Global Average Pool [6]. The 13,323 normalized facial images at a size of 224 x 224 that they preprocessed from the LFW dataset of nearly 5748 identities, are then aligned to generate a 128-dimensional matrix of each image as output from the neural network [8]. The preprocessing consists of detecting the face region in a complex background using AdaBoost [9]. Afterward, these preprocessed images are fed into a Resnet50 network to extract face and gender features through convolution layers. Before the output of these final images, the Global Average Pool is used to reduce the size of the network features. After this batch training, classification accuracy for each module was found: 99.33% for the face recognition module and 93% for the gender recognition module. The problem we have with this approach is that the classification accuracy could be further improved before feeding the dataset into the network, the authors would crop these images. Omkar et al. used a first 8-layer Alexnet neural network to perform a manual filter of 2.6 million images and obtained 982,803 images [10]. After training, the images obtained in the second 37-layer ConvNet neural network incorporating Dropout was 98.95% with the LFW dataset. Their proposed neural network uses dropout to avoid overlearning, but unfortunately, it lowers the performance of the model. The presence of false positives in the facial images may not cause an unregistered person in the system to be recognized. Mondal et al. used the GoogleNet neural network, which is composed of 22 layers integrating the triple loss function [11] . They extracted a 200-dimensional matrix on 13323 images of 5749 identities from the LFW dataset introduced in this neural network. The model was tested on 100 images of 10 people obtaining an accuracy of 99.1% when classified with a support vector machine. They proposed a system to secure votes, however, in case of bad votes, such as identical voters, although measures of similarity of the characteristics between the two faces of these voters can be too high, one of the two unregistered persons ends up voting and his record is deleted in the database. When the second person arrives to vote, he/she will not be able to (false positive). In this study, we use CNN to minimize the false positives of facial recognition to optimize its efficiency in use for multiple vote resolution. In order to verify the reliability of our system, we simulated an election where we undertook to add an already available dataset (LFW) to which we added the images of people taken on the internet without forgetting a small number of images taken by our voting system during 4 months (July-November 2021). This data will be used in our three polling stations (Ngousso, Koumassi, Biyemassi).

3 METHODOLOGY

As shown in Figure 1, our proposed facial recognition-based voting system consists of two main successive phases: the enrolment phase and the authentication phase. When the system captures an input facial image of each voter

using deep learning approach, it checks it with the features of the facial images that have been stored in the facial database. A feature similarity measure helps to compute the similarity comparisons so that the most dominant labels are attached as image labels. If the value of the similarity measured is less than a certain value, the classification result returns an error result, i.e. the system cannot identify the person. This is how facial recognition works, if two people are identical, their features are too similar [11]. Currently, face recognition usually consists of four modules: facial image acquisition, data preprocessing and augmentation, feature extraction with Resnet50, and feature comparison. Furthermore, a facial database is used to store face feature vectors and to perform feature comparisons.

3.1 Facial image collection

First, the OPENCV method operated under python, is used to collect the 150x150px facial images [12]. Images from the LFW dataset, images of people taken from the internet, and images taken by the OPENCV method will be used to verify the reliability of our system.

3.2 Preprocessing and data augmentation

In the preprocessing, the HOG method is used to detect the regions of the face, then a face landmark estimation algorithm is used to align the faces that are not centered. The idea of this algorithm is to find 68 landmarks that are present on each face. This algorithm detects the faces in a given image and then returns the landmarks of the face, thus the face shape. Finally, these images are cropped to a 47x55 px image by using OPENCV method. Data augmentation is done only during the first phase of enrolment [12]. We proceed to the data augmentation when there are few images using a machine learning algorithm. We have written two functions. The first one rotates the image takes as input the original image, the angle the coefficient, and returns as output the rotated image. The second function increases the data. This function takes the dataset as input and returns a rotated image as output. It starts by browsing the given set and retrieving the path to an image file. Then copies the image and applies the rotation function to it ten times, saving after each step. And finally, repeat the previous steps until all the images in the dataset have been traversed. For our case, we propose to increase horizontally the LFW dataset and thus it goes from 13 233 images to 173820 images.

3.3 Feature extraction

The feature extraction, 128-dimensional encoding values of the introduced voter face images, is done with the ResNet50 model described in Figure 2. The novelty of our model is that we have replaced the GAP layers with flatten layers and a succession of dense layers. First, we use the flatten layers to flatten the tensor in order to reshape to a shape equal to the number of elements contained in the tensor. These flatten layers receive as input the

output of the function named Resnet50 [13]. Then we use dense layers of 1024 and 512 parameters to further reduce the dimensions of the tensor. The dense 1024-parameter layers are first introduced just after the flatten layer, and then we insert the dense 512-parameter layers just below it. All these layers take as argument a glorot_uniform kernel initializer initialized in Keras, which is a relu activation. We initialize the initializer to produce the same random values over multiple calls [14] . And finally, we use the last dense layer to indicate the probabilities of each layer. This layer is placed below the last layers. It takes as an argument the number of classes corresponding to our classification problem, a softmax activation, and also a kernel initializer glorot_uniform initialized in Keras.

Mathematical formula :

Let X be the initial image input,
Let F(x) be the transformation function performed by ResNet50,
Let Y1 be the output after the flatten layer,
Let Y2 be the output after the dense layer (1024) with ReLU activation,
Let Y3 be the output after the dense layer (512) with ReLU activation,
Let Y4 be the final output after the dense layer (5745) with softmax activation.

The mathematical formula for your new model can be expressed as follows:

$$Y1 = \text{flatten}(F(X)) \tag{1}$$
$$Y2 = \text{relu}(Y1 * W2 + b2) \tag{2}$$
$$Y3 = \text{relu}(Y2 * W3 + b3) \tag{3}$$
$$Y4 = \text{softmax}(Y3 * W4 + b4) \tag{4}$$
$$\tag{5}$$

where flatten is the function that transforms the spatial dimensions of the input into a one-dimensional vector,
relu is the ReLU activation function,
softmax is the softmax activation function,
W2, W3, W4 are the weight matrices corresponding to each dense layer, and b2, b3, b4 are the corresponding bias vectors.

The formula indicates that output Y1 is obtained by applying the flatten layer to the output of ResNet50. Next, outputs Y1, Y2, Y3 and Y4 are calculated using the dense layers with corresponding weights and biases, and applying the ReLU activation function to outputs Y2 and Y3 and the softmax activation function to output Y4.

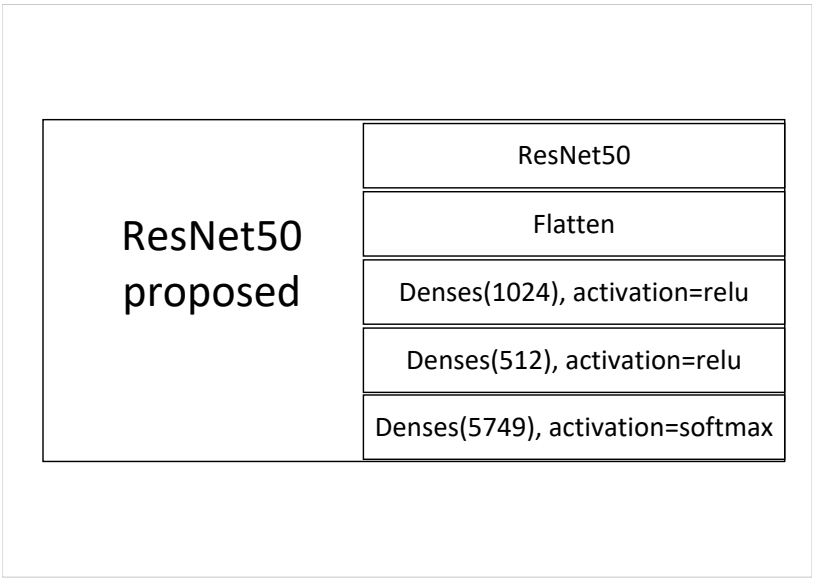

Fig. 2: Modified ResNet50 architecture

3.4 Facial database

The images, The RGB face images, information, and voter characteristics templates are stored in a relational database on a remote server.

3.5 Comparison of characteristics

The comparison of features is done through search and decision-making.

1. Search The search is performed from the 128 representations of the face of each person obtained from the extraction of the characteristics; while being based on the Euclidean distance corresponding directly to a measure of the similarity of the faces. These values are then compared to make a decision.
2. Decision The decision is made when the search finds a match or not. When it does match, the input image is labeled with its input name; this means that the similarity measure is above the threshold. In our context, the system will have the identity of this voter, and from a boolean variable, it tells if he/she has already voted or not. Otherwise, when this measure is below a certain threshold, the input image is labeled as unknown. This means that he/she will not be able to vote.

4 RESULTS AND INTERPRETATION

The facial recognition-based voting system was evaluated on a personal dataset and an LFW dataset.

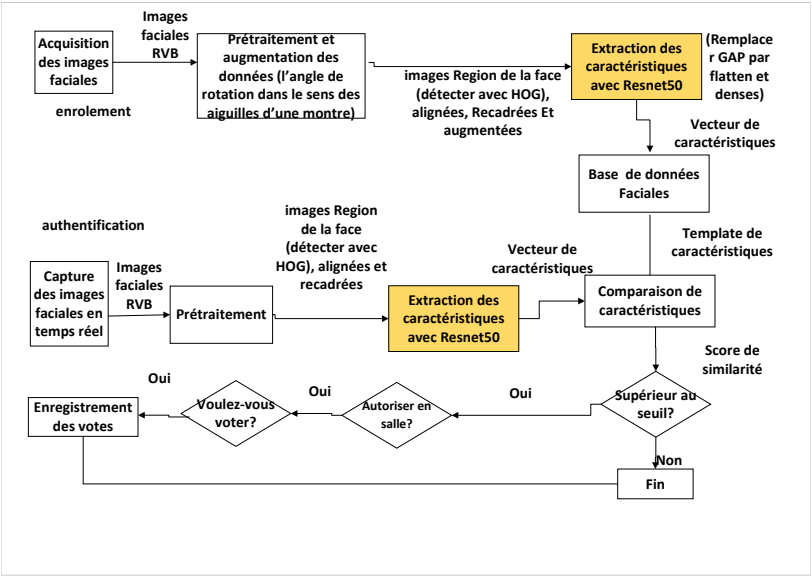

Fig. 3: A voting system based on facial recognition

4.1 Evaluation Metrics

The classifier is evaluated on classification rate, f-measure, recall and test precision.

$$Accuracy = \frac{VP + VN}{VP + FP + VN + FN}$$

$$recall = \frac{VP}{VP + FN}$$

$$precision = \frac{VP}{VP + FP}$$

Where TP, TN, FP, and FN represent respectively:

- True positives (TP): enrolled individuals (are recognized), can vote.
- True negatives (TN): non-enrolled individuals are not recognized to vote;
- False positives (FP): non-enrolled individuals are recognized to vote;
- False negatives (FN): enrolled individuals are not recognized to vote.

4.2 Datasets

The LFW dataset contains 5749 identities constituting 13323 images of RGB type and size 250x250 px. This dataset is then cropped to 47x55 px size and then augmented by the data augmentation algorithm which now contains 173820 images as mentioned above. This makes it contain 139069 images for training, 34751 images for validation and 34751 images for testing. LFW can be downloaded here [8]. The personal dataset contains 1170 identities constituting 13435 RGB images. It is also cropped to a size of 47 x55 px, so that it contains 10756 images for training, 2679 images for validation and 2679 images for testing.

4.3 Facial recognition performance

We trained the model used for face recognition in the LFW and personal datasets. From Table I, it can be seen that the results of each training vary across time periods. In the LFW dataset, with 10 batches of training, we obtained 99.51% validation classification rate, 99.92% validation test precision rate, 99.27% validation recall rate, and 0.0296 validation loss rate. After this dataset was trained on 20 batches, we obtained on the one hand, a slight improvement on these metrics including a validation classification rate of 99.60%, a validation recall rate of 99.50%, a validation loss rate of 0.0210, and on the other hand a slight decrease (the validation precision rate reduced from 99.92% to 99.84%). This means that on the one hand the more we increase the epochs, the more the model learns, the more the losses are reduced, but that it becomes saturated at a validation classification rate of 99.60% and on the other hand the validation precision rate does not increase with increasing batches but decreases slightly. The fact that the validation classification rate starts at 0.9 is explained by the fact that we proceeded just after cropping the images, with a machine-learning technique that increases the LFW data sets. This increase produces an average validation accuracy of 99.56% with an average validation loss rate of 2.19% when training the networks for 20 batches.

Table 1: 1st results

Dataset	Epoch	Loss	Acc	Val_loss	Val_acc	Precision	Val_Prec
lfw	10	0.0613	98.78%	0.0296	99.51%	99.80%	99.92%
	20	0.0081	99.86%	0.0210	99.60%	99.93%	99.84%
own	10	0.2296	97.11%	0.1469	98.51%	99.91%	100%
	20	0.0243	99.87%	0.0755	98.73%	99.99%	99.81%

We compared the previously reported results with those computed by our facial recognition voting system. As shown in Table 1 and 2, the comparison of the final results is performed in the same dataset. There is a clear noticeable improvement in the correction rate in our work.

Table 2: 2nd results

Dataset	Epoch	Recall	Val_Recall	Train Time
lfw	10	98.83%	99.27%	20 363s
	20	99.74%	99.50%	34 922s
own	10	93.22%	96.45%	1383s
	20	99.58%	98.17%	2794s

Table 3: ACCURACY OF CLASSIFICATION, TRAINING AND VALIDATION TEST, LOSS OF FACE RECOGNITION ON LFW DATASETS, PERSONAL

References	Datasets	Accuracy
Gumani et al. [15]	LFW	91.8%
Schroff et al [16]	LFW	98.87%
Omkar et al [10]	LFW	98.95%
Mondal et al [6]	LFW	99.1%
Yuxiang et al [6]	LFW	99.33%
This report	**LFW**	**99.56%**

4.4 Voting System Performance

After the votes are cast, the information of the voters who participated in the vote is generated in an Excel file. This file certifies that a voter has only had to vote once. In summary, we had a total of 102 people on the voters' list. 78 people showed up in this polling station, of which 72 people voted and 6 did not. Among the absentees, we counted 24, Figure 3.

5 CONCLUSION

The facial recognition voting authentication system combines in a way the voting system and the facial recognition system. The authentication system was proposed to solve the problems of multiple voting that could occur during the careful identification of voters simply because the system implemented does not control the fact that a voter can vote multiple times (vote in several different polling stations). This system was deployed based on ResNet50 neural networks applied to the Adam optimization algorithm, with a 10-5 learning rate. We used a flatten layer and a succession of dense layers instead of the global average pool (GAP) to improve the classification accuracy. This improvement may have been further refined due to the increase in cropped images we made. In this study, we conducted experiments on the different datasets and obtained a result. This result is compared with the results of the previous studies on the same dataset. After this comparison, it is found that we have obtained a better result i.e. 99.56

campaign	(Tous)				
Gender	(Tous)				
Number of voters	**Column labels**				
Labels lines	**(empty)**	**Yes**	**No**	**Absent**	**Grand total**
(empty)	**0**	**0**	**0**	**24**	**24**
(vide)	0	0	0	6	6
Ngousso	0	0	0	11	11
Biyemassi	0	0	0	7	7
13/11/2021	**0**	**56**	**6**	**0**	**62**
(vide)	0	4	0	0	4
Ngousso	0	35	4	0	39
Koumassi	0	4	0	0	4
Biyemassi	0	13	2	0	15
11/11/2021	**0**	**15**	**0**	**0**	**15**
Ngousso	0	12	0	0	12
Koumassi	0	2	0	0	2
Biyemassi	0	1	0	0	1
10/11/2021	**0**	**1**	**0**	**0**	**1**
Ngousso	0	1	0	0	1
Grand total	**0**	**72**	**6**	**24**	**102**

Fig. 4: summary of votes

6 FUTURE WORKS

The proposed system could be further improved by the following perspectives:

- The use of the triple loss function could further reduce the loss percentage of our model;
- The integration of fingerprint biometrics to further improve security;
- Also the use of blockchain to secure storage and exchanges on the cloud.

7 Declaration

7.1 Availability of supporting data

Supporting data is available on request.

7.2 Competing interests

The authors declare that they have no competing interests.

7.3 Funding

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

7.4 Authors' contributions

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the authors have contributed equally and that the authors order listed in the manuscript has been approved by all of us.

7.5 Acknowledgements

Not Applicable

References

[1] Stiftung, F.E.: Prévenir et lutter contre la fraude électorale au Cameroun. Yaoundé: Edition Clé (2012)

[2] Perronnin, F., Dugelay, J.-L.: Introduction à la biométrie-authentification des individus par traitement audio-vidéo. Traitement du signal **19**(4) (2002)

[3] Demers-Labrousse, N., Vandal, G., Aoun, S.: La Démocratie en Afrïque Subsaharienne: Le Cas du Cameroun. Université de Sherbrooke, ??? (2012)

[4] Kindzeka, M.E.: Claiming Massive Fraud, Cameroon Opposition Challenges Ruling Party Landslide Victory. https://www.voanews.com/a/africa_claiming-massive-fraud-cameroon-opposition-challenges-ruling-party-landslide-victor 6184185.html (2020)

[5] NSONGAN, P.M.: Cameroon bishops warn against election fraud. https://www.africanews.com/2018/10/11/cameroon-bishops-warn-against-election-fraud/ (2018)

[6] Zhou, Y., Ni, H., Ren, F., Kang, X.: Face and gender recognition system based on convolutional neural networks. In: 2019 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 1091–1095 (2019). IEEE

[7] Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: CVPR 2011, pp. 529–534 (2011). IEEE

[8] Huang, G.: Labeled Faces in the Wild Home. http://vis-www.cs.umass.edu/lfw/ (2018)

[9] Delbiaggio, N.: A comparison of facial recognition's algorithms (2017)

[10] Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition (2015)

[11] Mondal, I., Chatterjee, S.: Secure and hassle-free evm through deep learning based face recognition. In: 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 109–113 (2019). IEEE

[12] OpenCV, L.: 3: Computer Vision in C++ with the OpenCV Library/Kaehler A., Bradski G. O'Reilly Media (2017)

[13] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

[14] kerasy: Layer weight initializers. https://keras.io/api/layers/initializers/. (2022)

[15] Gurnani, A., Shah, K., Gajjar, V., Mavani, V., Khandhediya, Y.: Saf-bage: Salient approach for facial soft-biometric classification-age, gender, and facial expression. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 839–847 (2019). IEEE

[16] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)