Monitoring Neurological Disorder Patients via Deep Learning Based Facial Expressions Analysis

Muhammad Munsif[1,3], Mohib Ullah[2], Bilal Ahmad[2], Muhammad Sajjad[2(✉)], and Faouzi Alaya Cheikh[2]

[1] Islamia College University Peshawar, 25000 Peshawar, Pakistan
[2] Norwegian University of Science and Technology, 2815 Gjøvik, Norway
muhammad.sajjad@ntnu.no
[3] Sejong University, Seoul 143–747, South Korea

Abstract. Facial expression (FE) is the most natural and convincing source to communicate human emotions, providing valuable insides to the observer while assessing the emotional incongruities. In health care, the FE of the patient (specifically of neurological disorders (NDs) such as Parkinson's, Stroke, and Alzheimer's) can assist the medical doctor in evaluating the physical condition of a patient, such as fatigue, pain, and sadness. ND patients are usually going through proper observation and clinical tests, which are invasive, expensive and time-consuming. In this paper, an automatic lightweight deep learning (DL) based FEs recognition framework is developed that can classify the facial expression of ND patients with 93% accuracy. Initially, raw images of FEs are acquired from publicly available datasets according to the patient's most common expressions, such as normal, happy, sad, and anger. The framework cropped images through a face detector, extract high-level facial features through the convolutional layers and fed them to the dense layers for classification. The trained model is exported to an android based environment over a smart device and evaluated for real-time performance. The qualitative and quantitative results are evaluated on a standard dataset named Karolinska directed emotional faces (KDEF). Promising results are obtained of various NDs patients with Parkinson, Stroke, and Alzheimer that show the effectiveness of the proposed model.

Keywords: Neurological disorder · Convolution neural network · Parkinson · Alzheimer · Emotion recognition

1 Introduction

Human facial expressions (FEs) play a significant role in human-to-human interactions and human behaviour analysis. According to Mehrabian et al. [1], for effective oral communication, body language, including FEs, contributes up to

Published by Springer Nature Switzerland AG 2022
I. Maglogiannis et al. (Eds.): AIAI 2022 Workshops, IFIP AICT 652, pp. 412–423, 2022.
https://doi.org/10.1007/978-3-031-08341-9_33

55% of total importance, while voice tons and words contribute 38% and 7% respectively. Apart from this, FEs reflects common symptoms of various medical conditions like NDs including Parkinson's [2], Stroke [3], Alzheimer, and Bell Palsy [4] diseases. Most of the time, medical experts diagnose patients with ND problems through strict overtime monitoring and various invasive and expensive medical tests, which can be challenging and painful [5]. Thus, developing an alternative, cost-effective and endurable system is essential. An automatic FEs recognition system can assist a doctor in evaluating the ND patients' overall behaviour. Such a system can efficiently differentiate and identify various FEs to identify patients' conditions (e.g., feeling well, bad, normal) associated with clinical-related FEs features. These FEs linked with clinical features can be combined with the diagnostic process as biomarkers to evaluate the performance of therapeutic response toward an ND patient.

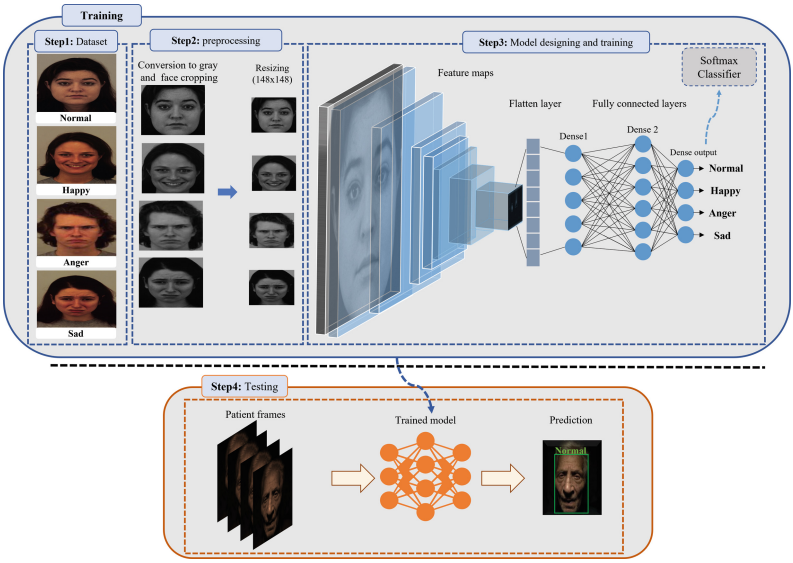

Fig. 1. Proposed framework, Broadly divided into two steps which include training and testing. Further, the training step consists of, dataset, pre-processing, and model designing, while the testing step contains real-time testing on real data of ND patients.

Various studies have been conducted to study the relationship between FEs and different NDs. Kohler et al. [6] conducted a study on Alzheimer patients' behaviour, and they found a deficit of FEs in Alzheimer patients. Similarly, Authors in [7] analyzed the behaviour of neurodegenerative disorder patients. They identified deficiency in most patients toward positive FEs such as happiness due to high subjectiveness to negative emotions such as anger and sadness. Ferndez et al. [8] observed impairment in the abilities of positive FEs recognition

in frontotemporal dementia and Alzheimer patients. To identify neurologically disordered by utilizing FEs, Authors in [9] developed a method to detect NDs using FEs. In their study, a photo/video containing different expressions is shown to the patients with NDs, and the patients are advised to mimic the expressions. The tool used in this study further decodes the expression of the patients by calculating the intensity of the imitated expression. Based on the calculated intensity, the system proposed in [9] predicts the state of the disease. In addition, Dantcheva et al. [10] proposed a computer vision-based framework to monitor severely demented people and their FEs during musical therapy, classifying activities and expressions during talking, singing, happy, and normal conditions. Similarly, authors in [11] proposed a machine learning (ML)based 3D mobile game application called JEMlmE to improve the expression skills of children with autism spectrum diseases. In this study, an ML model is trained over children's expressions (sadness, happiness, anger, and natural, etc.) and is integrated with JEMlmE. Playing JEMLmE, children produce different expressions and certain positive points through correct expressions, otherwise negative points. Further, Jin et al. [12] performed a comparative study of deep learning (DL) and ML-based techniques, diagnosing Parkinson's patients through FEs analysis. The authors collected videos of healthy and Parkinson patients containing smiley faces in this study. In Face++ API, traditional ML (such as SVM, DT, LR, RF) and DL based sequence learning (such as RNN and LSTM) are used for preprocessing, feature extraction and classification.

Apart from this, various FEs recognition techniques are developed to improve the performance of FEs recognition methods such as [13]. Among them Liang et al. [14] developed an action unit-based network to recognize 33 various fine-grained FEs. Similarly, in [15] authors proposed a generative adversarial network (GAN) based technique to solve the problem of bad artefacts while transforming one FE to another FE, for instance, sad to happy. Further, the adaptive learning-based FEs representation technique was proposed in [16] where authors developed a knowledgeable teacher and self-taught student network to learn facial emotions in both easy and complex environments adaptively. In addition, a cloud-based convolution neural network (CNN) framework was developed to recognize FEs recognition over edge server [17], Where the system captures a face image using a smartphone, transmitting it to the server for preprocessing and classification. State of the arts (SOTA) discussed high computational resources for training, testing, and deployment. The FEs in SOTA are not explicitly associated with the facial emotions of NDs patients for diagnostic purposes, only focusing on security and data quality applications. To cope with the critical challenges of computation, accuracy, and association of FEs with NDs patients for diagnostic purposes, we proposed a lightweight FEs recognition framework to assist the medical experts in early diagnosing of NDs patients. In a nutshell, the contributions of the proposed framework are three folds:

1. Developed a DL-based FEs analysis framework that can monitor early-stage NDs patients, including Parkinson's, Alzheimer's, and stroke patients.

2. The model of only 9 MB is achieved which is deployable in resource-constrained devices such as smartphones and tablets for the practical use of medical practitioners.
3. Achieved the highest accuracy of the model on NDs patients data collected from YouTube containing faces carrying numerous expressions belonging to different gender and age.

The rest of the paper is organized in the following order. Section 2 presents the data preparation step, including the details of the dataset, data pre-processing and augmentation. The model architecture and the training strategy is elaborated in Sect. 3. The experimental setup and the implementation details are given in Sect. 4. The quantitative results and ablation study is also presented in Sect. 4. Section 5 concludes the paper and gives potentials future research directions.

2 Data Preparation

Dataset collection, annotation, and arranging, especially in the case of FEs of ND patients, is a very challenging task. It requires a large number of patients suffering from ND or special skilled professional actors that can make a genuine expression like the ND patients. Both cases require substantial financial resources and substantial human efforts from the researcher, doctors, and patients. So instead of making a dataset from scratch, we have explored various publicly available datasets like the Japanese female facial expressions database JAFF [18], and KEDF [19]. Further details of dataset and its preparation for the DL model are listed below.

2.1 Dataset

KEDF is a publicly available dataset developed by the psychological section of the department of clinical neuroscience, Karolinska Institute, Sweden. It contains universal human facial expressions (Normal, happy, sad, surprised, afraid, angry, and disgusted) images having the size of 562×762 of 70 participants (35 males and 35 females) obtained from five different angles with various cameras. We selected the KEDF dataset for the training of the proposed model because it contains clear, varied, and high-resolution images. Further, in Neurological disorders, patients mainly express four expressions: normal, happy, anger, and sad. So, we chose only these classes of data from the KDEF and arranged them in four classes as shown in step 1 of Fig. 1 accordingly. The arranged data consists of 900 RGB images in each of the four classes split between the training and validation set. Due to this split, 80% of the data is used for training and 20% for evaluation. Further, for real-time testing on real patients' we collected a full-length video from the YouTube platform for each mentioned NDs patient by searching in different well-known channels like Michigan Medicine, 60 min Australia BAYSTATEHEATH. After collection, we extract frames from each video and select frames or parts of the video to pass from the trained model for real-time evaluation based on the expression and age of the patients.

Table 1. Hyper-parameters of the proposed model

Layer	Kernel size	No of kernels/Neurons	Activation	Dropout rate(%)
Conv2d_1	3×3	32	Relu	–
B-norm1	–	–	–	–
Conv2d_2	3×3	32	Relu	–
B-norm2	–	–	–	–
Max-pool1	2×2	–	–	–
Conv2d_3	3×3	64	Relu	–
B-norm3	–	–	–	–
Max-pool2	2×2	–	–	–
Conv2d_4	3×3	64	Relu	–
B-norm4	–	–	–	–
Max-pool3	2×2	–	–	–
Conv2d_5	5×5	128	Relu	–
B-norm5	–	–	–	–
Max-pool4	2×2	–	–	–
Conv2d_6	5×5	128	Relu	–
B-norm6	–	–	–	–
Dropout1	–	–	–	–
Max-pool5	2×2	–	–	–
Flatten	-	–	–	–
Dropout2	–	–	–	30
Dense1	–	64	Relu	–
B-norm7	–	–	–	–
Dense2	–	64	Relu	–
Dropout3	–	–	–	30
Output Dense	–	4	Softmax	–

2.2 Pre-processing

Preprocessing is one of the critical steps to improve the learning capabilities of the model during training. Preprocessing aims to remove unessential pixels from the raw images and keep only region of interest (ROI) for processing. The first step is to detect the face and then crop it, as shown in Fig. 2. Face detection is a challenging task due to angles and illumination variations. To avoid such variations, a popular algorithm in terms of accuracy for face detection called viola jones [20] is used. RGB images are converted to grey before feeding them to the viola jones algorithm. Further, to reduce the computational cost, the cropped images are downsampled to 148×148 before feeding them into the proposed training model.

Fig. 2. Face detection and cropping

3 Model Architecture

In order to design an efficient DL model that is easily deployable on resource-constrained devices such as smartphones, it is essential to have a minimal number of trainable and non-trainable parameters. These parameters are directly related to the different components of the model and its hyper-parameters. The broad graphical depiction of our proposed model is given in Fig. 1. It consists of various components, including convolution, pooling, batch normalization, dropout, and dense layers. The model accepts a grayscale image of 148×148 as input and provides predicted probabilities as output for four facial expressions categories. The architecture contains six convolutions layers (CLs) with various numbers of 3×3 and 5×5 filters in the first four and last two layers, respectively. Relu activation function is used in each CL, which helps the model avoid high vanishing gradient problems and learn complex nonlinear functions while training. Five max-pooling layers (MPL) are utilized with the kernel size of 2×2 after each CL except the first one to reduce the dimensions of resulting features maps from CLs and leave only high weighted features as output. Further, seven batch normalization layers (BNLs) are kept after each layer for standardization of the input batches for CLs and to smooth convergence during the model's training. In last, two hidden layers have 64 neurons with the relu activation function in each, and one dense output layer contains four neurons and a SoftMax activation function used to acquire probability for four classes as an output of the model. Besides this, the dropout @ of 30% regularization technique is utilized before each of the last three dense layers to avoid overfitting and achieve high accuracy on validation samples. Further, the final model contains a total of 1.3 million parameters. A visual view of our proposed model is shown in Fig 1 and hyper-parameters parameters of various layers are tabulated in Table 1.

4 Experiment

In this section, we present the evaluation performance of the proposed method. First, we explain the experimental setting, then datasets used in the model's training and evaluation, followed by evaluation metrics, ablation study, and real-time testing. All these steps are discussed below in detail.

4.1 Experimental Setup and Implementation Details

The implementation and experiments were carried out in python version 3.7 based virtual environment that is installed on a personal computer with the specification of GTX GeForce 1070 GPU, intel(R) Xeon(R) X5560 processor with 2.80 GH clock speed, and Install memory (RAM) 8.00 Giga bite. Further, different frameworks and libraries are utilized, including TensorFlow-GPU version 2.0.0 with the frontend of Keras-GPU for designing, training, and evaluating the DL model. Categorical cross-entropy loss function and Adam optimizer with an initial learning rate of 10^{-4} are used to calculate the loss of the model and update its weights while training, respectively. In addition, we trained the proposed model on 32 minibatch sizes for 150 epochs which took almost one and half hours. Apart from this, NumPy is used for various mathematical operations like reshaping and concatenation, and Matplotlib is utilized to visualise different evaluation graphs.

Fig. 3. Real-time testing, the first row is of the Parkinson, second is Alzheimer and the last one shows the result of our framework on stroke patients

4.2 Evaluation Metrics

A total of six matrices are used to evaluate the performance of models. The confusion matrix is shown in Table 3. Time inference is used to check the model's speed and evaluate the model's weight after training. Model loss is used to show the model performance verification during training. In addition, for a better and more accurate comparison, all these metrics are calculated using the Keres functions.

4.3 Confusion Matrix

For better evaluation and observing the class-wise performance of the proposed, we draw the confusion matrix of the model, which is depicted as a Fig. 4. It can be observed that the performance of the model is 96% for each happy and neutral. However, performance for the Angry and sad class is low, which is 88% and 89% respectively.

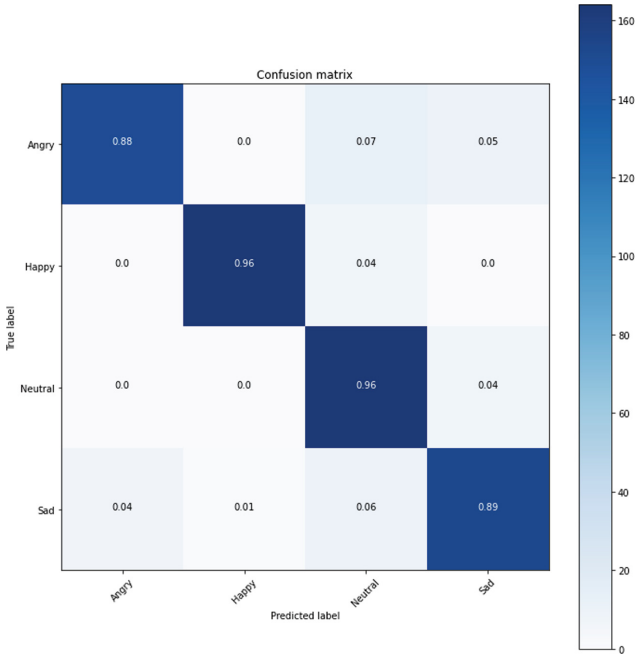

Fig. 4. Confusion matrix of proposed FER model

4.4 Ablation Study

We have done experiments on two different datasets. First, the model is trained and evaluated on KEDF data. Secondly, the real-time evolution of the trained model on ND patients' data was collected from the YouTube platform. The results of these experiments are tabulated in Table 2. From experiments, we found that when the number of convolution layers increases, dense layers are kept constant (two), training and validation accuracies are improved gradually due to the learning capability of high-level and accurate face features extraction. However, the model's size increased due to the number of trainable parameters in convolution layers. Further, when we added more dense layers, the model becomes overfit because of the high complexity in the last layers. An example of

this case is experiment 3 in Table 2, where the addition of another dense layer found the model highly overfit. Despite this, from top to bottom in Table 2, we can observe an increase in inference time. This is due to increases in features extraction layers where each CNN layer takes a specific amount of time. As a result of these extensive experiments, we achieved a high performer model having only six convolutions and two dense layers with 96.0% training and 97.0 % validation accuracies, and 0.25 training and 0.18 validation losses. In the end, the total size of the proposed model was achieved at only 9 Mbs which can be easily deployed on resource-constrained devices. Further, for better evaluation, Fig. 5 depicts the performance of the proposed model for 150 epochs where a gradual increase in accuracy and decrease in loss can be observed throughout the training session.

Table 2. Experiments using different variant of our proposed DL architectures

Conv layers	Dense layers	Train accuracy(%)	Validation accuracy(%)	Inference time (sec)
1	2	85	77	0.199
2	2	87	82	0.299
3	3	90	68	0.392
4	2	92	86	0.554
5	2	94	92	0.749
6	2	96	97	0.852

4.5 Real-time Evaluation

The real-time testing result is shown in Fig. 3. The first row indicates our model's performance on images of the early-stage Parkinson's patient getting treatment from the doctors. The rest of the two are early-stage Alzheimer's and Stroke patients, respectively. Further, all patient's expressions are recognized correctly. However, certain difficult situations are wrongly classified. For example, Alzheimer's patient is normal in actuality but classified as angry due to a very drastic change in angle and appearance of the patient face.

Table 3. Comparison with state-of-the-art methods

Model name	Testing accuracy (%)	Recall(%)	Precision(%)	F1 Score(%)
NASNet mobile	74	74	74	72
ResNet50	80	80	80	80
Mobile NetV2	73	73	73	72
Proposed	93	93	93	93

4.6 Performance Comparison with State-of the Art Models

For performance comparison with the existing state of the art model, we used a pre-trained model, including NasNet mobile, Mobile Net V2, and ResNet50. The performance of each model is tabulated as Table 3. The ResNet achieved 80% testing accuracy, recall, precision and F1 score. On the other hand, the proposed model achieved the highest 93% accuracy for each mentioned metric.

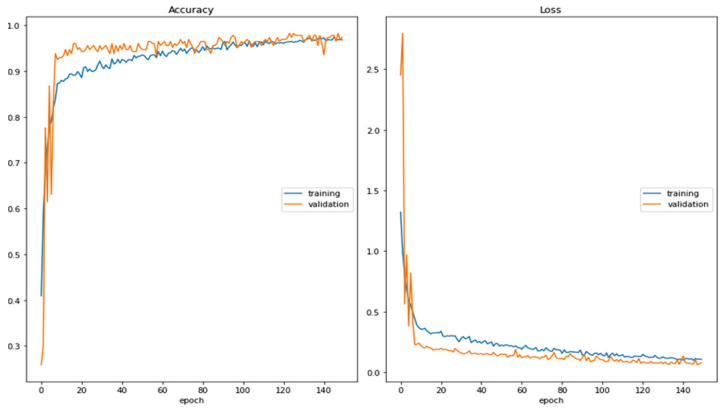

Fig. 5. Accuracies and losses of the proposed model

5 Conclusion

We presented a DL based system for automatic FEs analysis of NDs such as Parkinson's, Alzheimer's, and stroke patients. After various experiments, we achieved a lightweight and accurate model having accuracy up to 96.0% of training and 97.0% of validation. Further, tested our model in real-time using the real data of NDs patients. Besides this, the system is able successfully deploy on resource-constrained devices due its lightweight. In the future, we plan to collect more challenging datasets of the patients and improve the system through attention mechanisms and incorporating temporal information with spatial information of specific FEs.

Acknowledgement. The European Union funded this research through the Horizon 2020 Research and Innovation Programme, in the context of the ALAMEDA (Bridging the Early Diagnosis and Treatment Gap of Brain Diseases via Smart, Connected, Proactive and Evidence-based Technological Interventions) project under grant agreement No GA 101017558.

References

1. Mehrabian, A.: Some referents and measures of nonverbal behavior. Behav. Res. Meth. Instrum. **1**(6), 203–207 (1968)
2. Ricciardi, L., et al.: Facial emotion recognition and expression in Parkinson's disease: an emotional mirror mechanism? PloS one **12**(1), e0169110 (2017)
3. Lin, J., Chen, Y., Wen, H., Yang, Z., Zeng, J.: Weakness of eye closure with central facial paralysis after unilateral hemispheric stroke predicts a worse outcome. J. Stroke Cerebrovasc. Dis. **26**(4), 834–841 (2017)
4. Baugh, R.F., et al.: Clinical practice guideline: bell's palsy. Otolaryngol.-Head Neck Surg. **149**(3_suppl), S1–S27 (2013)
5. Chen, X., Wang, Z., Cheikh, F.A., Ullah, M.: 3D-resnet fused attention for autism spectrum disorder classification. In: International Conference on Image and Graphics, pp. 607–617. Springer (2021)
6. Kohler, C.G., et al.: Emotion-discrimination deficits in mild Alzheimer disease. Am. J. Geriatr. Psychiatry **13**(11), 926–933 (2005)
7. Mandal, M.K., Pandey, R., Prasad, A.B.: Facial expressions of emotions and schizophrenia: a review. Schizophrenia Bull. **24**(3), 399–412 (1998)
8. Fernandez-Duque, D., Black, S.E.: Impaired recognition of negative facial emotions in patients with frontotemporal dementia. Neuropsychologia **43**(11), 1673–1687 (2005)
9. Bevilacqua, V., D'Ambruoso, D., Mandolino, G., Suma, M.: A new tool to support diagnosis of neurological disorders by means of facial expressions. In: 2011 IEEE International Symposium on Medical Measurements and Applications, pp. 544–549. IEEE (2011)
10. Dantcheva, A., Bilinski, P., Nguyen, H.T., Broutart, J.C., Bremond, F.: Expression recognition for severely demented patients in music reminiscence-therapy. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 783–787. IEEE (2017)
11. Dapogny, A., et al.: Jemime: a serious game to teach children with ASD how to adequately produce facial expressions. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 723–730. IEEE (2018)
12. Jin, B., Yue, Q., Zhang, L., Gao, Z.: Diagnosing Parkinson disease through facial expression recognition: video analysis. J. Med. Internet Res. **22**(7), e18697 (2020)
13. Alreshidi, A., Ullah, M.: Facial emotion recognition using hybrid features. In: Informatics, vol. 7, p. 6. Multidisciplinary Digital Publishing Institute (2020)
14. Liang, L., Lang, C., Li, Y., Feng, S., Zhao, J.: Fine-grained facial expression recognition in the wild. IEEE Trans. Inform. Forens. Secur.**16**, 482–494 (2020)
15. Wu, R., Zhang, G., Lu, S., Chen, T.: Cascade ef-gan: Progressive facial expression editing with local focuses. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5021–5030 (2020)
16. Li, H., Wang, N., Ding, X., Yang, X., Gao, X.: Adaptively learning facial expression representation via CF labels and distillation. IEEE Trans. Image Process. **30**, 2016–2028 (2021)
17. Shirian, A., Tripathi, S., Guha, T.: Dynamic emotion modeling with learnable graphs and graph inception network. IEEE Trans. Multimedia (2021)
18. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. IEEE (1998)

19. Lundqvist, D., Flykt, A., Öhman, A.: The karolinska directed emotional faces (kdef). CD ROM Depart. Clin. Neurosci. Psychol. Sect. Karolinska Institutet **91**(630), 2–2 (1998)
20. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, p. 1. IEEE (2001)