Rehabilitation of Traumatic Brain Injured Patients

Patient Mood Analysis from Multimodal Video

Ilyas, Chaudhary Muhammad Aqdus; Nasrollahi, Kamal; Rehm, Matthias; Moeslund, Thomas B.

Published in:
2018 IEEE International Conference on Image Processing

DOI (link to publication from Publisher):
10.1109/ICIP.2018.8451223

Publication date:
2018

Rehabilitation of Traumatic Brain Injured Patients: Patient Mood Analysis from Multimodal Video

Chaudhary Muhammad Aqdus Ilyas*, Kamal Nasrollahi*, Matthias Rehm† and Thomas B. Moeslund*
*Visual Analysis of People Lab †Robotics Aalborg U
Aalborg University, Denmark
Email: {cmai,kn, matthias,tbm }@create.aau.dk

***Abstract*—Rehabilitation after traumatic brain injury (TBI) is very critical as it is largely unpredictable depending upon the nature of the injury. Rehabilitation process and recovery time also varies, as it takes months and years, depending upon the assessment of treatment, mental and physical conditions and strategies. Due to non-cooperative behaviour of patients, and increase in negative emotional expressions it is very beneficial to evaluate these expressions in a contactless way, and perform a rehabilitation physiotherapy, cognitive or other behavioral activities when the patient is in a positive mood. In this paper we have analyzed the methods for facial features extraction for TBI patients to determine optimal time to have aforementioned rehabilitation process on the basis of positive and negative facial expressions. We have employed a deep learning architecture based on convolutional neural network and long short term memory on RGB and thermal data that were collected in challenging scenarios from real patients. It automatically identifies the patient's facial expressions, and inform experts or trainers that "it is the time" to start rehabilitation session.**

I. Introduction

Traumatic brain injury (TBI) causes life-long damage to cognitive, physical, behavioural and social functions. It may take up to 5 years or more for recovery after TBI [1]. According to International Brain Injury Association (IBIA), annually one million people suffer from traumatic brain injury (TBI) only in America whereas same number of people suffer with TBI in Europe [2]. American Center for Disease Control and Prevention estimates more than 3.7 million people are living with long term disability after TBI [3]. During rehabilitation period, patient has to live in a specialized care center called neuro-center or care home where the main focus is on the retraining of activities of daily life, cognitive, social and physical exercises through a set of protocols. Recovery targets are based on determination of combination of cognitive, behavioral and physical shortfalls. It is seen that rehabilitation activities are performed daily on set time table of neuro-center, regardless of mental conditions of subject. This leads to more time expensive training with less result oriented outcome.

There is high urgency of fast and accurate rehabilitation process so the TBI patients have to spend less time in care centers or have to suffer less with limited independence and low quality of life. Caregivers, trainers or experts dealing with TBI patients face severe difficulty in performing rehabilitation activities as the patients have limited or reduced ability to perceive social and interaction signals [4]. In addition to that there is relative increase in negative emotions like depression, anger, anxiety, sadness, verbal or physical aggression and lack of social communication after TBI [5][6]. Extra consideration and care need to be made while interacting with these patients. Experts and trainers believe that with assessment of impact of injury to positive and negative emotions, caregivers can provide more accurate and faster rehabilitation services[7]. Goal and activity setting, for brain injury rehabilitation by involving patients emotional states, increase the chances of faster recovery with broader aspects[6]. It will provide flexibility to staff to work around with many more patients at the same neuro-center in less time.

Experts are putting emphasis on implementing Computer Vision (CV) techniques in health care sector as population is growing, so as the number of brain injured patients. Therefore, automatic diagnosis of mental and physical health states through unobtrusive computer vision techniques by using facial features has rapidly increased since past decades [8][9][10]. The fundamental approach for utilizing these CV techniques is to diminish the errors by human assessment. Furthermore, these approaches are cost effective as compared to medical examination by physicians or doctors, and can provide continues monitoring of the patients.

Existing CV techniques for facial expression recognition (FER) systems are mostly designed and implemented for healthy people. However, TBI patients' emotional states are quite different from healthy people as they have high degree of imbalance of six common emotional expressions accompanied by reduced muscle movement or paralysis. The database established for TBI patients for FER described in our previous paper [5], shows that it is very difficult to have all six expressions. Therefore, in this paper we suggest to classify the facial features into two emotional states either positive and negative. If patients are found to be in a positive mood, the caregivers are alarmed to start the rehabilitation. Furthermore, we do bimodal analysis of facial images in both the color RGB and thermal modalities. To do this, we have expanded our previous database of [5] by including more TBI patients. Experts and psychologists have been asked to help us annotating the collected data. They characterized positive expression as smile, laugh, surprise and few unique neutral expressions, while fear, disgust, anger, sad, stress and fatigue are categorized as negative expressions, sometimes additionally associated with lips trembling, teeth grinding and frequent eye blinking[11]. In case of TBI patients, negative expressions are more frequent as compared to positive ones. Our obtained experimental results

using deep learning techniques show that the two employed modalities can complement each other on classifying patients status to positive or negative.

In terms of methodology, contributions by [12] and [13] are probably most close to our method but these systems work well for healthy people in controlled environment. Moreover these systems have luxury of data sets where subjects were cooperative with no or less pose variation, minimum occlusions and high quality images unlike with TBI patients. As described in our previous paper [5], our database in [5] was established with Face Quality Assessment (FQA) but with only contained RGB images. In the current paper, we have improved the database with both RGB and thermal images with additional subjects and more pre-processing techniques like face frontalisation. We have verified the proposed system with real data of TBI patients collected in real environment at neuro-center where these TBI individuals are looked after 24/7.

The rest of this paper is organized as follows: The related work on FER are reviewed in the next section. Section III describes the new database including data collection and pre-processing techniques. Section IV describes the proposed methodology for facial feature extraction and expression recognition. Section V presents the results obtained from the experiments. Finally, Section VI concludes the paper.

II. Related Work

Current FER system can be categorized on the basis of methods used for feature extraction and classification. Our main focus is on the methods involving Convolution Neural Networks (CNN) or other deep learning approaches as they provide state of the art results for, e.g., face recognition [14] [15] [16], facial expressions recognition [17] [18] [19] [20] [21] [22] [23] [12] [13] and emotional states identification [24] [25] [26] [27]. Handcrafted features such as Local Binary Pattern (LBP), SIFT, Local Quantized Pattern (LPQ) and Histogram of Oriented Gradients (HOG) applied in [28] [29][30][31][32] are outperformed by CNN based deep neural networks despite their low computational cost.

In [17], Tang proposed deep CNN along with Support Vector Machines (SVM) and achieved state of the arts results for FER with 1st prize in FER-2013 competition. In 2014, Liu [19] performed three functions- feature learning, feature selection and classification in unified manner through Boosted Deep Belief Networks (BDBN). This method worked exceptionally well even for extremely complicated features from facial image. [22] used DBN models to overcome the limitations of linear feature selections. Yu and Zang [20] in 2015, presented their work for Emotion recognition in Wild challenge for image based static FER. They have applied multiple deep CNN with random initialization of each network and minimized likelihood and hinge loss. Their results surpassed the challenge baseline significantly. In year 2017, [13] exercised CNN to learn features from VGG-Faces and integrated with Long Short Term Memory (LSTM) to gain the temporal information. This approach was further improved by [12] who applied deep CNN for features classification into expressions and feed the system with super-resolved facial images.

III. TBI Patient Database for FER

A. Data Acquisition

To analyze facial expressions, data is collected in three pre-specified scenarios from seven TBI patients in two modalities: RGB and Thermal. Pre-specified scenarios in data collection are maintained to have reliable data for further use. Those scenarios are: 1) cognitive activity 2) physiotherapy and 3) social communication. These scenarios are selected after consulting many experts and care givers, who are working on rehabilitation of TBI individuals in Denmark. On contrary to healthy people, as mentioned in [5], data acquisition task is quite complicated due to extreme behavioural responses, verbalization, physical aggression, impaired reasoning, reduced cognitive skills along with frequent pose variations.

Ilyas et al. [5], collected RGB database by Axis RGB-Q16 camera with resolution of 1280 x 960 to 160 x 90 pixels at 30fps (frames per second) and applied pre processing techniques of face detection, FQA, (Supervised Decent Method) SDM for landmark detection and tracking before logging into a face log. We have operated with a Logitech camera as well to record the starting and ending time stamp of particular expressions. Along with RGB, we have gathered thermal images of TBI subjects with Axis Thermal-Q1922 camera with focal lens of 10 mm. RGB cameras are prone to difficulties in challenging conditions like shadows or when subject are obscured with complex background. Thermal cameras, on the other hand, can provide addition information of a scene. Thermal and RGB imagery are synchronized with the help of time stamps and annotation are made in sequence of facial expressions. Both RGB and thermal images are collected with same 30 fps. Furthermore, homography estimation is employed for image registration by determining homography matrices from RGB to thermal by [33].

TABLE I
Database Of TBI patients with Activity Participation

Subjects	Number of Sessions	Activities Participated		
		Cognitive	Social Comm	Physiotherapy
Subject A	7	Y	Y	Y
Subject B	5	Y	Y	Y
Subject C	5	Y	Y	Y
Subject D	7	Y	X	Y
Subject E	3	Y	X	X
Subject F	4	Y	Y	Y
Subject G	3	X	Y	Y

B. Database Structure

Data is collected from seven TBI patients in 34 sessions on the above mentioned three pre-specified scenarios. Few subjects did not take part in all activities, details are described in Table I. Two categories of expressions are recorded: Positive Expression (PE) and Negative Expressions (NE). PEs are smile, laugh, surprise and few unique neutral expressions, while NEs are fear, disgust, anger, sad, stress and fatigue. We

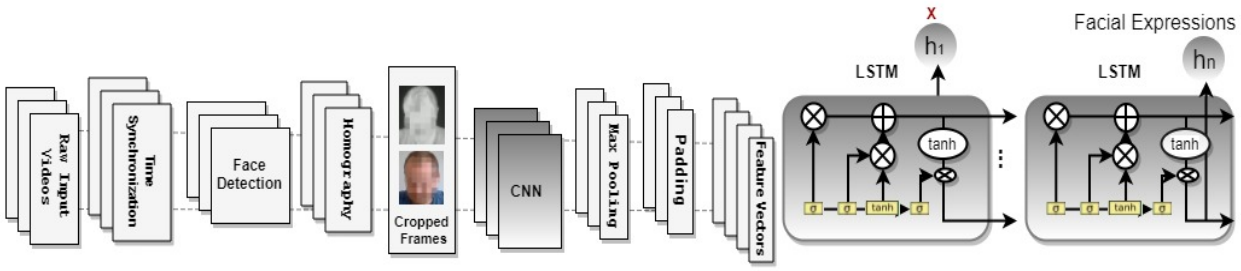

Fig. 1. CNN+LSTM based deep learning architecture for both modalities to exploit spatio-temporal information for FER.

have got 861 video events, each of maximum 5 seconds in length.

IV. The Proposed Methodology

This section presents the architecture of the intended approach for FER analysis of real TBI patient in realistic environment. We have employed the same method as followed in [5] but employed new pre-processing technique of face frontalization because of large pose variation. We tested the deep learning method of [5] on both modalities, with early and late fusions. Facial expressions are recognized by employing CNN (to use spatial features) and linking with LSTM to utilize spatio-temporal attributes of RGB, thermal and fused RGB-thermal modalities. The block diagram of the proposed method is illustrated in Figure 1. The steps of the proposed system are further explained in the following subsections.

A. Pre-Processing

Firstly, the face is detected, and facial landmarks are identified and tracked using [34] from a synchronized input video. TBI patients have large pose variations so to avoid loss of information, the posed faces are rotated using a frontalization algorithm. For face frontalization, landmarks are calculated with arbitrary facial positions and by finding inverse of the transpose matrix, the face is frontalized. In next step, face cropping is done in RGB modality, and associated faces in thermal modality is cropped by applying a homography. Homography is a special technique that allows geometric transformation of fixed points from one plane to another. In this case, RGB and thermal planes are homo-graphed with subject face. To remove erroneous detection and ensuring high quality of images, face quality assessment is applied before feeding the faces into the CNN pipeline.

B. CNN + LSTM Architecture

After the pre-processing of the data, it is fed to 2D-CNN for training purpose for mood recognition based on PE and NE. This network is fine tuned by VGG-16 face model [35] for spatial feature extraction. CNN parameters are initialized randomly and through back propagation using gradient descent its weights are adjusted. Thermal data is also fine tuned with pre-trained VGG-16 face (RGB) model. CNN deals with frames in isolated manner. For capitalizing on relation with time, special Recurrent Neural Network (RNN) called LSTM is employed. LSTM is gate controlled network with input (i), output(o) and forget (f) gates. LSTM gates holds the input information as long as its forget gate is not triggered to acquire the temporal information between frames for said purposes. These gates control the flow of instructions by point wise multiplication and sigmoid functions σ, which bound the information flow between zero and one by the followings:

$$i(t) = \sigma(W_{(x\to i)}x(t) + W_{(h\to i)}h(t-1) + b_{(1\to i)}) \quad (1)$$

$$f(t) = \sigma(W_{(x\to f)}x(t) + W(h\to f)h(t-1) + b_{(1\to f)}) \quad (2)$$

In these equations, W are weights associated with activated neurons for particular input i. Where as σ squashes the value of activation between the range of 0 and 1

$$z(t) = tanh(W_{(x\to c)}x(t)) + W_{(h\to c)}h(t-1) + b_{(1\to c)}) \quad (3)$$

$$c(t) = f(t)c(t-1) + i(t)z(t), \quad (4)$$

$$o(t) = \sigma(W_{(x\to o)}x(t) + W_{(h\to o)}h(t-1) + b_{(1\to o)}) \quad (5)$$

$$h(t) = o(t)tanh(c(t)), \quad (6)$$

where $z(t)$ is the input to the cell at time t, c is the cell, and h is the output. $W_{(x\to y)}$ are the weights from x to y. In the classification, LSTM finally provides a decision score for the expression recognition.

C. Fusion Scheme

In order to analyze the ability of both modalities in FER applications, two approaches were employed: 1) data level fusion (early) 2) feature level Fusion. In the first approach both modalities are combined into data array for feature learning through CNN. In the second method, both RGB and thermal imagery features are fed separately into deep learning system for feature learning and combined together as input for second classifier (LSTM) for final output. Block diagram of both modalities can be seen in Figure 2.

V. Experimental Results

We demonstrate the results in the following contexts:

a) Classification of six basic expression groups in both early and feature level fusion scenarios to evaluate the performance of CNN+LSTM based FER

b) PE and NE classifications before and after face frontalization on all individual modalities and fusions.

TABLE II
RECOGNITION ACCURACY OF PROPOSED METHOD IN DIFFERENT CONTEXTS

Confusion Matrix %	RGB Non-Frontal		RGB Frontal		Thermal		Early Fusion		[13]		Feature Level Fusion	
	PE	NE	PE	NE	PE	NE	PE	NE	PE	NE	PE	NE
Positive Expression (PE)	0.75	0.17	0.86	0.15	0.69	0.25	0.84	0.14	0.79	0.12	0.86	0.11
Negative Expression (NE)	0.21	0.71	0.11	0.87	0.21	0.65	0.16	0.79	0.1	0.82	0.09	0.89
Recognition Accuracy (%)	79.34		86.93		74.45		84.39		87.97		89.74	

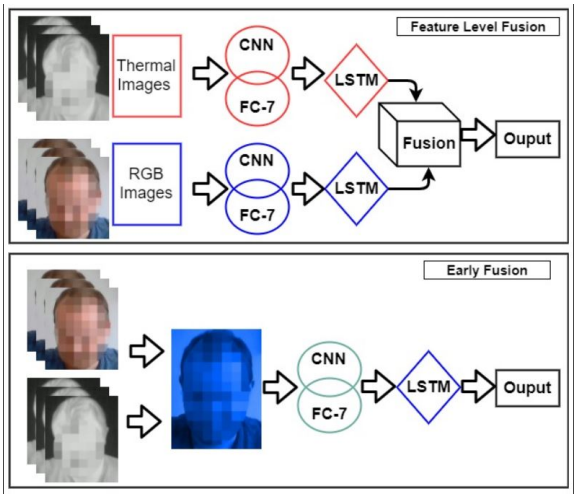

Fig. 2. Block diagram of early and Feature Level Fusion of modalities for FER.

TABLE III
CONFUSION MATRIX BY EARLY FUSION OF MODALITIES FOR 6 BASIC FER.

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	0.77	0.03	0.02	0.07	0.07	0.01
Happy	0.04	0.71	0.02	0.03	0.05	0.16
Angry	0.04	0.02	0.81	0.09	0.03	0.02
Sad	0.07	0.01	0.05	0.76	0.13	0.01
Fatigued	0.09	0.01	0.09	0.1	0.55	0.11
Surprised	0.07	0.14	0.1	0.02	0.06	0.56

First we produced results of positive and negative mood identification (based on PE and NE) without employing face frontalization (FF) and then with face frontalization. It is seen in table II column 1-4, after FF recognition accuracy is increased to 86.93 percentage from 79.34. In second case, we trained our system for thermal data, true positive and true negative are 69 and 65 percentage with high miss classification rate of 23.74 percentage. Overall recognition accuracy is achieved up to 74.45 percentage. In next stage we combined both RGB with FF to thermal data in early fusion scheme and obtained accuracy of 84.39 percentage for mood recognition. We also employed early and feature level fusion to analyze the results for 6 common facial expressions in Table III and Table IV. In both cases, fatigue and surprise have less recognition

TABLE IV
MCONFUSION MATRIX BY FEATURE LEVEL FUSION OF MODALITIES FOR 6 BASIC FER.

	Neutral	Happy	Angry	Sad	Fatigued	Surprised
Neutral	0.77	0.03	0.02	0.07	0.07	0.01
Happy	0.04	0.71	0.02	0.03	0.05	0.16
Angry	0.04	0.02	0.81	0.09	0.03	0.02
Sad	0.07	0.01	0.05	0.76	0.13	0.01
Fatigued	0.09	0.01	0.09	0.1	0.55	0.11
Surprised	0.07	0.14	0.1	0.02	0.06	0.56

accuracy due to less available data. If we compare table II with table III and IV, we can see that accuracy of system is increased for positive and negative expressions as compared to all 6 expressions. In the next stage we employed the [13] system on our database II. It is observed that its accuracy is 87.97 percentage much lesser than [13] 97.2 percentage, when he implemented on CK+ database. In last stage, we employed the feature level fusion and achieved 89.74 percentage of accuracy. By feature level fusion, despite computational expensive surpassed other state of art methods for positive and negative expression recognition. That shows that our system is producing competitive results with challenging data sets.

VI. CONCLUSIONS

Mood recognition is important task for rehabilitation and care centers. In this work we have faced the challenge of mood recognition of TBI patients rather than facial expression recognition for healthy people. In case of TBI individuals, extraction of all expression is very complicated and its dependant to patient disability and FER did not provide good results [5]. However, we recognized the mood of patients with accuracy of 86.93 percentage that is very close to [13] system when implemented on TBI patient database. So this system can help physiotherapist and trainers in fast rehabilitation process after recognizing the positive mood of the patient. Furthermore, we applied early and feature level fusion to enhance the recognition rate of the system. Our system results can be improved further by employing 3D face frontalization. Even though the results are encouraging, efforts are still in progress to provide the robust solutions to deal with real time and environment challenges like real time computation or patient positioning.

REFERENCES

[1] I. J. B. Fary Khan and I. D. cameron, "Rehabilitation after brain injury," *The medical Journal of Australia*, vol. 178, pp. 290–295, March 2003.

[2] Brain injury facts—international brain injury association-ibia. [Online]. Available: http://www.internationalbrain.org/brain-injury-facts/

[3] T. CA, B. JM, B. MJ, and X. L., "Traumatic brain injuryrelated emergency department visits, hospitalizations, and deaths united states, 2007 and 2013," *Morbidity and Mortality Weekly Report (MMWR)*, vol. 66(No. SS-9), p. 116, 2017.

[4] J. Bird and R. Parente, *Recognition of nonverbal communication of emotion after traumatic brain injury*, 2014.

[5] C. M. A. Ilyas, M. A. Haque, M. Rehm, K. Nasrollahi, and T. B. Moeslund, "Facial expression recognition for traumatic brain injured patients," in *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,*, INSTICC. SciTePress, 2018, pp. 522–530.

[6] B. Dang, W. Chen, W. He, and G. Chen, "Rehabilitation treatment and progress of traumatic brain injury dysfunction," *Neural Plasticity*, vol. 2017, 2017.

[7] A. Bender, C. Adrion, L. Fischer, M. Huber, K. Jawny, A. Straube, and U. Mansmann, "Long-term rehabilitation in patients with acquired brain injury," *Deutsches rzteblatt International*, vol. 113, pp. 634–641, September 2016.

[8] F. Li, C. Zhao, Z. Xia, Y. Wang, X. Zhou, and G.-Z. Li, "Computer-assisted lip diagnosis on traditional chinese medicine using multi-class support vector machines," *BMC Complementary and Alternative Medicine*, vol. 12, no. 1, p. 127, Aug 2012.

[9] M. P. Hyett, G. B. Parker, and A. Dhall, *The Utility of Facial Analysis Algorithms in Detecting Melancholia*. Cham: Springer International Publishing, 2016, pp. 359–375.

[10] Y. Chen, *Face Perception in Schizophrenia Spectrum Disorders: Interface Between Cognitive and Social Cognitive Functioning*. Dordrecht: Springer Netherlands, 2011, pp. 111–120.

[11] K. Lander and S. Metcalfe, "The influence of positive and negative facial expressions on face familiarity," *Memory*, vol. 15, no. 1, pp. 63–69, 2007, pMID: 17479925. [Online]. Available: https://doi.org/10.1080/09658210601108732

[12] M. Bellantonio, M. A. Haque, P. Rodriguez, K. Nasrollahi, T. Telve, S. Escalera, J. Gonzalez, T. B. Moeslund, P. Rasti, and G. Anbarjafari, *Spatio-temporal Pain Recognition in CNN-Based Super-Resolved Facial Images*. Cham: Springer International Publishing, 2017, pp. 151–162.

[13] P. Rodriguez, G. Cucurull, J. Gonzlez, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. Roca, "Deep pain: Exploiting long short-term memory networks for facial expression classification," *IEEE Transactions on Cybernetics*, vol. PP, no. 99, pp. 1–11, 2017.

[14] H. Li and G. Hua, "Hierarchical-pep model for real-world face recognition," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 4055–4064.

[15] Z. Huang, R. Wang, S. Shan, and X. Chen, "Face recognition on large-scale video in the wild with hybrid euclidean-and-riemannian metric learning," *Pattern Recognition*, vol. 48, no. 10, pp. 3113 – 3124, 2015, discriminative Feature Learning from Big Data for Visual Recognition. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320315001120

[16] J. Yang, P. Ren, D. Chen, F. Wen, H. Li, and G. Hua, "Neural aggregation network for video face recognition," *CoRR*, vol. abs/1603.05474, 2016. [Online]. Available: http://arxiv.org/abs/1603.05474

[17] Y. Tang, "Deep learning using support vector machines," *CoRR*, vol. abs/1306.0239, 2013. [Online]. Available: http://arxiv.org/abs/1306.0239

[18] S. E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, c. Gülçehre, R. Memisevic, P. Vincent, A. Courville, Y. Bengio, R. C. Ferrari, M. Mirza, S. Jean, P.-L. Carrier, Y. Dauphin, N. Boulanger-Lewandowski, A. Aggarwal, J. Zumer, P. Lamblin, J.-P. Raymond, G. Desjardins, R. Pascanu, D. Warde-Farley, A. Torabi, A. Sharma, E. Bengio, M. Côté, K. R. Konda, and Z. Wu, "Combining modality specific deep neural networks for emotion recognition in video," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, ser. ICMI '13. New York, NY, USA: ACM, 2013, pp. 543–550. [Online]. Available: http://doi.acm.org/10.1145/2522848.2531745

[19] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1805–1812.

[20] Z. Yu and C. Zhang, "Image based static facial expression recognition with multiple deep network learning," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI '15. New York, NY, USA: ACM, 2015, pp. 435–442. [Online]. Available: http://doi.acm.org/10.1145/2818346.2830595

[21] M. Liu, S. Li, S. Shan, R. Wang, and X. Chen, "Deeply learning deformable facial action parts model for dynamic expression analysis," in *Computer Vision – ACCV 2014*, D. Cremers, I. Reid, H. Saito, and M.-H. Yang, Eds. Cham: Springer International Publishing, 2015, pp. 143–157.

[22] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *Journal on Multimodal User Interfaces*, vol. 10, no. 2, pp. 173–189, Jun 2016. [Online]. Available: https://doi.org/10.1007/s12193-015-0209-0

[23] I. Ofodile, K. Kulkarni, C. A. Corneanu, S. Escalera, X. Baró, S. J. Hyniewska, J. Allik, and G. Anbarjafari, "Automatic recognition of deceptive facial expressions of emotion," *CoRR*, vol. abs/1707.04061, 2017. [Online]. Available: http://arxiv.org/abs/1707.04061

[24] M. Liu, R. Wang, S. Li, S. Shan, Z. Huang, and X. Chen, "Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild," in *Proceedings of the 16th International Conference on Multimodal Interaction*, ser. ICMI '14. New York, NY, USA: ACM, 2014, pp. 494–501. [Online]. Available: http://doi.acm.org/10.1145/2663204.2666274

[25] Y. Fan, X. Lu, D. Li, and Y. Liu, "Video-based emotion recognition using cnn-rnn and c3d hybrid networks," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ser. ICMI 2016. New York, NY, USA: ACM, 2016, pp. 445–450. [Online]. Available: http://doi.acm.org/10.1145/2993148.2997632

[26] H. Ranganathan, S. Chakraborty, and S. Panchanathan, "Multimodal emotion recognition using deep learning architectures," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2016, pp. 1–9.

[27] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari, "Audio-visual emotion recognition in video clips," *IEEE Transactions on Affective Computing*, vol. PP, no. 99, pp. 1–1, 2017.

[28] M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, vol. 5, pp. 4525–4536, 2017.

[29] X. Zhao and S. Zhang, "Facial expression recognition based on local binary patterns and kernel discriminant isomap," *Sensors*, vol. 11, no. 10, pp. 9573–9588, 2011.

[30] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hogebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537 – 1543, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167865508001104

[31] S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, "3d facial expression recognition using sift descriptors of automatically detected keypoints," *The Visual Computer*, vol. 27, no. 11, p. 1021, Jun 2011. [Online]. Available: https://doi.org/10.1007/s00371-011-0611-x

[32] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803 – 816, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885608001844

[33] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[34] X. Xiong and F. D. la Torre, "Supervised descent method and its applications to face alignment," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 532–539.

[35] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.