

# Face Analysis in the Wild

Flávio H. de B. Zavan, Nathaly Gasparin, Júlio C. Batista, Luan P. e Silva,  
Vítor Albiero, Olga R. P. Bellon and Luciano Silva

IMAGO Research Group

Universidade Federal do Paraná (UFPR)

Av. Francisco H. dos Santos, 100, 81531-980, Curitiba, PR, Brazil

E-mail: {flavio,nathaly.gasparin,julio.batista,luan.porfirio,valbiero,olga,luciano}@ufpr.br

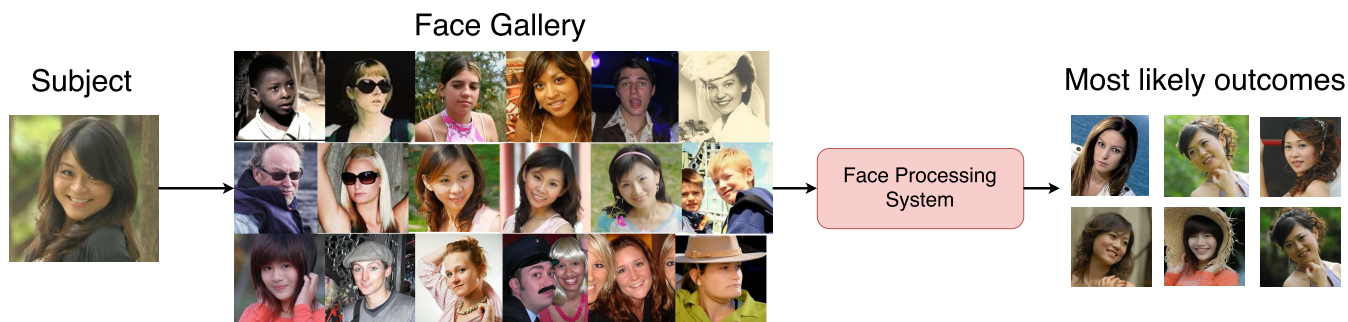


Fig. 1: Challenging search in an in the wild dataset [1]

**Abstract**—With the global demand for extra security systems, and the growing of human-machine interaction, facial analysis in unconstrained environments (in the wild) became a hot-topic in recent computer vision research. Unconstrained environments include surveillance footage, social media photos and live broadcasts. This type of images and videos include no control over illumination, position, size, occlusion, and facial expressions. Successful facial processing methods for controlled scenarios are unable to pledge with challenging circumstances. Consequently, methods tailored for handling those situations are indispensable for the face analysis research progress. This work presents a comprehensive review of state-of-the-art methods, drawing attention to the complications derived from in the wild scenarios and the behavior differences when applied to the controlled images. The main topics to be covered are: (1) face detection; (2) facial image quality; (3) head pose estimation; (4) face alignment; (5) 3D face reconstruction; (6) gender and age estimation; (7) facial expressions and emotions; and (8) face recognition. Finally, available code and applications for in the wild face analysis are presented, followed by a discussion on future directions.

## I. INTRODUCTION

Recent research on facial image analysis has shifted its focus towards performing face recognition and expression identification in unconstrained environments (in the wild) [2]–[4], including surveillance footage, photos posted on social media, and live broadcasts. In the wild images are characterized by varying head poses, positions, and size; cluttered backgrounds; variations and non-uniformity in illumination; facial expressions, including open and closed mouths; and occlusion caused by accessories and other objects (Fig. 1).

The lack of control over the image’s capture sensor carries out problems that require preprocessing steps to be smoothed out. An example is the use of generic quality measures to evaluate image conditions or face hallucination, a technique that synthesizes high resolution images from their low resolution form. Traditional methods that achieve high performance when processing images acquired in controlled scenarios are unable to yield substantial, useful results in more challenging situations. Therefore, methodologies specifically tailored for handling such scenarios are necessary in order to overcome these limitations.

While holistic face processing solutions exist [5], [6], most published works treat each subproblem separately. It is possible to study and evaluate each challenge in isolation, however, a combination of these is typically applied as a pipeline for ultimately performing face recognition or facial expression analysis. An overview of the major face processing steps is presented in Fig. 2.

Face detection, locating the faces in a given image, is essential for many tasks in face processing. Head pose estimation determine the head orientation relative to the camera, and is important to a variety of face processing approaches. Another crucial task for many applications is face alignment, which determines the geometric structure of the face. It is directly applied to 3D face reconstruction, the recovery of the shape and appearance of the face, which can assist face recognition, as it is invariant to scale and robust to occlusion. Facial expression analysis emerged from the necessity for systems

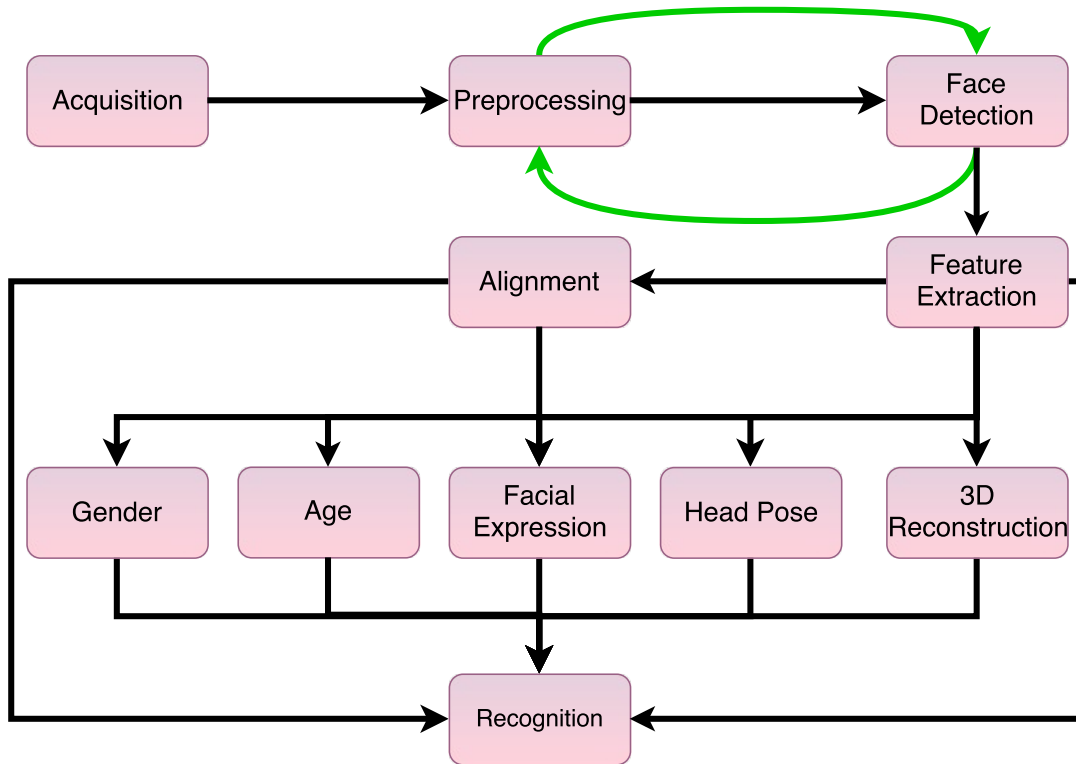


Fig. 2: Overview of a face processing system with possible workflows. Face detection may occur before preprocessing

to interact and comprehend humans, and is the estimation of muscle movements patterns that translates into known expressions of emotions.

A comprehensive presentation, contextualization and review of the different steps is presented, focusing on the final recognition and expression analysis applications. While the order in which processing occurs is not fixed, topics are presented in a common, logical order, minimizing logical conflicts.

This paper addresses face detection in Section II; face quality in Section III; head pose estimation in Section III; face alignment in Section V; 3D reconstruction in Section VI; gender and age estimation in Section VII; facial expression analysis in Section VIII; face recognition in Section IX; datasets in Section X; code availability in Section XI; to conclude, Section XII presents final remarks.

## II. FACE DETECTION

Face detection is defined as determining the position and size of a face in an image. The traditional solution, based on cascading Haar features, proposed by Viola and Jones [7] is shown to perform adequately in controlled scenarios with limited variations in lightning, expressions and head poses. However, when applied in challenging situations, its performance degrades significantly [5], [8], [9].

A more recent method, proposed by Liao *et al.* [9], is an evolution of the Viola-Jones methodology [7]. A cascade of Normalized Pixel Difference (NPD) features is used for achieving a reliable and fast face detection approach (Fig. 3).

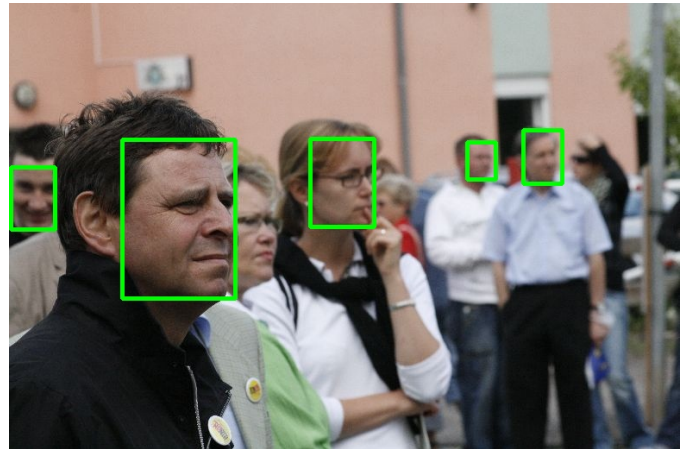


Fig. 3: Example face detection using Faster R-CNN [10] on an image from the AFLW dataset [11]

The state-of-the-art for face detection in the wild was established by the use of CNNs. Faster R-CNN [10] is a generic object detector based on the introduced concept of Region Proposal Networks (RPN). The RPN are used to generate region proposals, which are evaluated by another network as an object or not.

Jiang and Learned-Miller [12] studied in depth its application for face detection, and determined a set of parameters for maximizing its performance on the task. YOLO [13] and SSD [14] are state-of-the-art real-time object detectors and,

similar to Faster R-CNN, can be adapted for face detection. Moreover, a scale invariant, CNN based face detector was proposed by Hu and Ramanan [15]. Alternatively, Ranjan *et al.* [5] proposed a complete framework for face processing in challenging scenarios, and is able to accurately detect the face while outperforming other existing methods. Faster R-CNN [10] and the work proposed by Hu and Ramanan [15] offer increased performance at the expense of greater computational cost. Alternatively, YOLO [13] and SSD [14] provide solutions for cases when real time processing is needed.




### III. FACIAL IMAGE QUALITY

Preprocessing of challenging images can be performed by using generic quality measures, such as the ones proposed by Abaza *et al.* [16] for evaluating face regions. However, due to differences in focus and blur that can be found in the wild images, it is common for the preprocessing step to take place after the face is detected (Fig. 2).

Dutta *et al.* [17] proposed a Bayesian model for predicting face recognition performance on images with varying illumination and head pose. Abaza *et al.* [16] generate a quality score by evaluating illumination, brightness, contrast, focus and sharpness of the face region. Silva *et al.* [18] extended this evaluation by including a sixth measure, a head pose score calculated given the nose region [19] (Table I).

The use of the head pose for estimating the face quality allows for estimates that closely relate to the final problem being solved. It is common for face recognition and expression analysis methods to achieve better performance on frontal faces. To favor these scenarios, this metric can be used to select frames from a video when temporal information is available.

TABLE I: Example quality measurements proposed by Abaza *et al.* [16] and the head pose [18]. Images were extracted and modified from a video in the 300-VW dataset [20]

			
<b>Contrast</b>	0.350713	0.127890	0.301992
<b>Brightness</b>	0.431940	0.393008	0.389695
<b>Focus</b>	0.034787	0.009951	0.006392
<b>Sharpness</b>	0.085802	0.026179	0.053955
<b>Illumination</b>	0.369130	0.347084	0.317291
<b>Head Yaw</b>	0°	-45°	-15°

A common challenge in unconstrained environments is the low resolution of the captured faces, which negatively affect face analysis performance [21]. Face hallucination was proposed to address this issue by generating a high resolution equivalent of the degraded face based on temporal information and previously seen examples [22]. Recently, this same technique has been applied for synthesizing sketches [23], however, focus is kept on the original application, sometimes referred to as super-resolution.

Jiang *et al.* [24] proposed an iterative approach for generating hallucinated faces, preserving the original high resolution geometry. Zhou *et al.* [21] proposed a learning approach by extracting robust face representations from the raw input using a CNN for generating high quality faces. Recently, the novel use of smooth regressions was proposed [25], constraining reconstruction while combining internal and external samples, achieving consistent results.

### IV. HEAD POSE ESTIMATION

Head pose estimation is defined by determining the angle of rotation of the head relative to the camera on at least one of three axis, yaw, pitch and roll [19] (Table I). While it is traditionally linked to gaze estimation [26], recent works have successfully used the head pose for estimating the face quality [18] and performing face alignment [27].

Initially, when studied on controlled environments, multiple solutions using traditional computer vision tools were proposed [26]. However, under challenging circumstances, robust approaches became necessary. Zhu and Ramanan [28] proposed a mixture of trees for performing face detection, landmark localization and head pose estimation simultaneously. Manifold analysis of the face region has also been applied for solving this problem [29], [30]. Recently, a complete face processing solution based on CNNs was proposed by Ranjan *et al.* [6], including head pose estimation. Zavan *et al.* [19] propose an alternative approach, in which only the nose region is used, as opposed to the whole face, allowing for consistent estimation even when the face is degraded.

### V. FACE ALIGNMENT

Face alignment is the task of identifying the geometric structure of faces on images using discriminant facial components, *i.e.* eye corners, nose and mouth [31]. Although it is common to use the set of 68 landmarks defined in [32], the number of landmarks might vary depending on the task. For example, [33] used five landmarks to rotate a facial image through an affine transformation, but [34] used 94 landmarks for facial expression analysis.

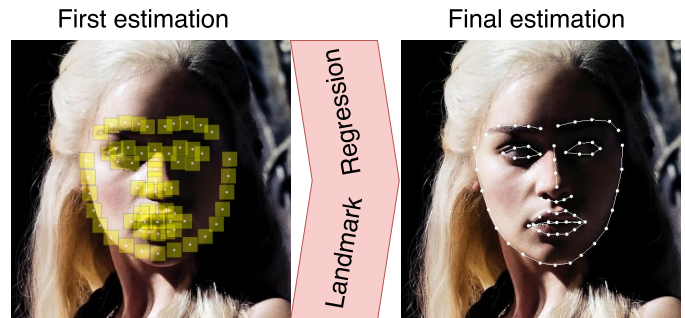


Fig. 4: Example face alignment cascade regression refinement [35]

Face alignment approaches include methods based on cascade regressions [36], [37], which are not robust to head pose variations. In order to address this issue, Zhu *et al.*

[38] and Yan *et al.* [39] used a range of facial geometries to initialize the regression model, while Yang *et al.* [40] combined pose information with a rough alignment. State-of-the-art performance was achieved by approaches based on CNNs. Sun *et al.* [33] used a CNN to localize five landmarks, and Trigeorgis *et al.* [41] proposed the usage of recurrent neural networks.

The presented methods are used only for near frontal faces. To handle this issue, face alignment was extended from 2D to 3D. One of the main efforts towards this change was the 3D Face Alignment in the Wild Challenge (3DFAW) [42], which evaluated pose invariant 3D landmark localization. The method proposed by Zavan *et al.* [27] detected landmarks by fitting a mean shape model according to the head pose, estimated using the nose region, and was extended by Silva [35] using cascade regressions (Fig. 4). Zhao *et al.* [43] proposed estimating 2D landmarks using a CNN, and then inferring the depth using a set of fully connected layers. Finally, the state-of-the-art challenge was achieved by Bulat and Tzimiropoulos [44] using a two-state convolutional part heatmap regression to locate 2D landmarks, and then estimating the depth using a residual network.

## VI. 3D RECONSTRUCTION

Humans perceive the 3D shape of objects and faces through patterns of shading and geometry [47]. Methods for 3D face reconstruction try to recover this information in 2D images or video frames. The uses for precise 3D face models range from security based systems capable of recognizing a subject independently of expression and facial accessories, to entertainment systems transporting someone's looks into a virtual world.

A 3D morphable model (3DMM) is a principal component analysis (PCA) built model of many face scans in dense correspondence [48], [49], and is one of the most studied approaches for 3D face reconstruction. The resultant 3D model can be morphed by estimating the principal components coefficients, following age, ethnicity and other human facial characteristics restrictions. The seminal work from Vetter and Blanz [50], later improved by Romdhani *et al.* [51] was developed in an environment with controlled illumination and background. Such scenarios are not suitable for practical applications. Problems found in the wild are usually related to background segmentation, landmark detection under large poses and illumination, and scale.

Shape from shading approaches can retrieve a face's 3D model directly from the image's shading information and have been tested in unconstrained scenarios [52]. Recently, Liang *et al.* [53] used images of the same person obtained on the internet for reconstructing the whole head. This approach makes use of different head poses from different images of the same subject.

Huber *et al.* proposed a landmark fitting based approach, which estimates the coefficients with the geometry of the face only, without using the texture. Piotraschke *et al.* [54] proposed a new approach for selecting the best out of many

in the wild images of the same person before reconstructing the face with a 3DMM. Tran *et al.* [45] proposed a novel approach, which uses a CNN for fitting the 3DMM. Results obtained using two approaches with available code [45], [46] are presented in Fig. 5.

## VII. GENDER AND AGE ESTIMATION

In computer vision, gender estimation is defined strictly as determining if a subject is male or female. Age estimation admits variations, as datasets may be annotated with the precise age [55] or a general age range [56]. Both tasks are typically solved using only the face.

It is common for approaches to provide a single solution for determining the gender and the age of a given subject in unconstrained environments. Tasks can be performed simultaneously [6], or the same architecture can be reused [57].

In recent work, Ranjan *et al.* [6] was able to accurately estimate the gender and age of subjects in challenging scenarios, while simultaneously performing other face analysis tasks. Levi and Hassner [57] proposed a CNN architecture that can be trained for both tasks, either classifying a face into two genders or eight age classes. Similarly, Mansanet *et al.* [58] provided a solution for estimating the gender by combining local features and CNNs. To estimate the age under considerable facial expressions, Lou *et al.* [59] proposed a joint-learning approach using a graphical model that learns the relationship between age and expression, achieving precise results.

## VIII. FACIAL EXPRESSIONS AND EMOTIONS

Facial expression identification is performed by using the Facial Action Coding System (FACS) proposed by Ekman *et al.* [60]. This model is based on the movement of facial muscles, known as Action Units (AUs). FACS is a categorical model stating that there is a limited set of emotion expressions [61], the basic six (happy, sad, angry, disgust, fear, surprise). Another model is the Circumplex of Affect [62], which uses a continuum of two dimensions based on valence (pleasant/unpleasant) and arousal (relaxed/aroused).

There is an ongoing debate on which one of the models best describes human understanding of facial expressions. However, it is known that there are more than six basic facial expressions [34], [63]. The main difference between FACS and the Circumplex of Affect is: FACS analyzes facial expressions that can be used to express emotions, while the Circumplex of Affect describes affect which can rely on facial expressions, but also on other signals such as audio and gestures.

Facial expression analysis and emotion recognition used to be evaluated on datasets acquired in controlled environments [64], [65]. These datasets are also known as posed, as the subjects were asked to perform the six basic expressions; resulting in clearly defined and intense renditions. However, daily expressions are subtler and not so distinct. To surpass this limitation, datasets of people showing spontaneous expressions were created [66]–[68]. The spontaneous behavior was recorded while subjects performed a task to elucidate specific

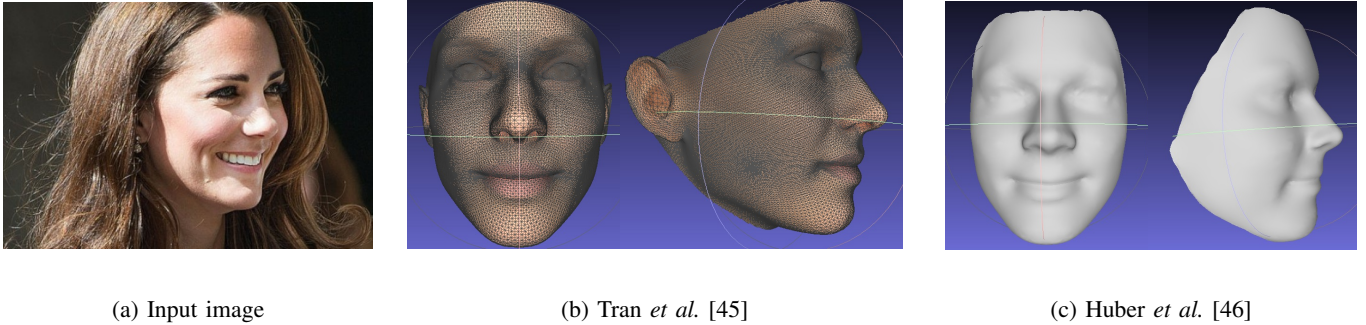


Fig. 5: Example of two approaches for 3D face reconstruction of an image from the 300-W dataset [32]

expressions. Although these datasets provide spontaneous expressions, they were also recorded in controlled environments.

Because facial expressions in the wild are complex, due to their spontaneous behavior, using a model with a fixed number of expressions has limited usage. Thus, performing AU detection and intensity estimation is more appropriate for this scenario. AU detection is a binary task of predicting if a given AU is present (active) or not. It is possible to use a binary model for each AU [4], [69], as it is also possible to model the task into a multilabel classification problem [70], [71]. Although it is possible to understand expressions by only detecting AUs, a better understanding can be achieved by estimating their intensities. AU intensity is important for estimating compound facial expressions, as some AUs interfere with each other [63], [72]. AU intensity estimation predicts the contraction level of the AU; it is only required when the AU is active. The contraction levels vary from A to E (or 1 to 5) with the middle intensities being the most common ones [60]. In this case, it is possible to use a multiclass [73] or a regression [69] model for each AU. Alternatively, a model can be used to perform a joint optimization for all AUs [70], [72], [74].

Although many approaches for AU detection and intensity estimation have been proposed, it is not clear how to handle these problems in challenging environments. One of the most recent approaches, based on Gabor filters, was successfully used to annotate a large dataset of facial expressions in the wild [4]. Open challenges for performing facial expression analysis in unconstrained environments remain [75]. An example is the head pose, which was shown to be one of the main issues in facial expression analysis [76]. Furthermore, the Facial Expression Recognition and Analysis Challenge (FERA) 2017 [77] focused on solving AU detection and intensity estimation under various head poses; the winners [78], [79] were CNN based approaches.

## IX. FACE RECOGNITION

Processing faces in unconstrained environments has direct applications in security systems, particularly for face recognition. Wang *et al.* [3] explore this aspect by performing large scale searches for individuals in datasets containing millions of subjects. They also perform a case study of the Boston

Marathon bombing, simulated using the suspects' pictures that were originally shared to match with their social media pictures enrolled in a large gallery of subjects.

Parkhi *et al.* [80] leveraged the availability of large face datasets by suggesting the use of very deep CNNs, which, when properly trained, are able to achieve state-of-the-art results for face recognition. To study the need for such large labeled datasets, Similarly, Amos *et al.* [81] proposed a CNN based face recognition method also optimized using a triplet loss of identities embeddings. Masi *et al.* [82] proposed manipulating the original training images by synthesizing new head poses, shapes and facial expressions through 3D reconstruction. Results indicate considerable gains in classification accuracy without the need for including new training images in the original dataset.

Focusing on in the wild images, Ranjan *et al.* [6] proposed a complete, all-in-one, CNN, including detection, alignment, gender, pose and age estimation, smile detection, and face recognition. Experiments achieve high accuracy for all tasks, which indicates that performing multiple tasks at once can be positive for such approaches as the initial layers of the CNN can extract features that are useful to all applications.

## X. DATASETS

An overview of the main benchmarks for each of the tasks related to face analysis is shown in Table II. It is important to note that, for some tasks, controlled datasets are used because, to the best of our knowledge, there are no in the wild labelled datasets for that task. This is the case for 3D reconstruction and AU intensity estimation. Although in the wild datasets exist for AU detection, it is common to evaluate the proposed approaches in controlled environments.

In order to simulate in the wild scenarios, for 3D reconstruction, it is a common practice to use pictures from the internet [45], [53], [102]; for facial expression analysis cross-dataset evaluation is being used, as described in [4].

## XI. CODE AVAILABILITY

The source code of many of the presented methods is available to the scientific community to be used for benchmarking and building upon. Table III presents an overview of all these methods.

TABLE II: Overview of the datasets for face analysis. The acronyms in the first column represent the target task of the dataset. The acronyms stand for: Face Detection (FD), Head Pose Estimation (HPE), 3D Reconstructions (3D), Face Alignment (FA), Gender and Age Estimation (GA), Emotion Recognition (EM), Facial Expression Analysis (FEA), and Face Recognition (FR)

	Dataset	Environment	Type	Description
FD	FDDDB [83]	in the wild	Image	5,000+ faces
	Wider Face [84]	in the wild	Image	300,000 faces
HPE	AFW [28]	in the wild	Image	Head pose from $-90^\circ$ to $90^\circ$ in steps of $15^\circ$
	AFLW [11]	in the wild	Image	Continuous annotation acquired through 3D model fitting
3D	Multi-PIE [85]	controlled	Image	15 view points and 19 illumination conditions
	Bosphorus [86]	controlled	Image	Includes the 3D ground truth for the faces
FA	300-VW [20]	in the wild	Video	68 landmarks per face
	3DAFW [42]	in the wild	Image	66 3D landmarks
	LS3D-W [87]	in the wild	Image	68 3D landmarks form 230,000 images
	Menpo [1]	in the wild	Image	29 or 68 landmarks depending on the head pose
GA	ChaLearn [88]	in the wild	Image	12,000 images
	CelebA [89]	in the wild	Image	10,000 subjects
EM	EmotioNet [4]	in the wild	Image	23 emotions
	AffectNet [90]	in the wild	Image	8 emotions and valence/arousal
	FER-Wild [91]	in the wild	Image	7 emotions
	FER-2013 [92]	in the wild	Image	7 emotions
	EmotiW [93]	in the wild	Image/Video	7 emotions
	AVEC [94]	controlled	Audio/Video	Valence/Arousal
	Aff-Wild [95]	in the wild	Images	Valence/Arousal
FEA	EmotioNet [4]	in the wild	Image	Occurrence of 12 AUs
	AM-FED [96]	in the wild	Image	Occurrence of 12 AUs
	FERA 2011 [97]	controlled	Image	Occurrence of 12 AUs
	FERA 2015 [98]	controlled	Video	Occurrence of 12 AUs; Intensity of 5 AUs
	FERA 2017 [77]	controlled	Video	Occurrence of 10 AUs; Intensity of 7 AUs; 9 head poses
	BP4D [67], [68]	controlled	Video	Occurrence of 12 AUs; Intensity of 5 AUs
	DISFA [66]	controlled	Video	Intensity of 12 AUs
FR	PaSC [99]	in the wild	Image/Video	265-293 identities
	LFW [100]	in the wild	Image	5,749 identities
	YTF [101]	in the wild	Video	1,595 identities
	Parkhi <i>et al.</i> [80]	in the wild	Image	2,600 identities

TABLE III: Available state-of-the-art code

Category	Method
Face Detection	[10] <sup>1</sup> [9] <sup>2</sup> [15] <sup>3</sup>
Head Pose Estimation	[28] <sup>4</sup>
Face Alignment	[87] <sup>5</sup>
3D Reconstruction	[46] <sup>6</sup> [45] <sup>7</sup>
Age and Gender Estimation	[57] <sup>8</sup>
Facial Expression Analysis	[69] <sup>9</sup> [71] <sup>10</sup> [74] <sup>11</sup> [72] <sup>12</sup>
Face Recognition	[80] <sup>13</sup> [69] <sup>14</sup>

<sup>1</sup><https://github.com/rbgirshick/py-faster-rcnn>

<sup>2</sup><http://www.cbsr.ia.ac.cn/users/scliao/projects/npdface/>

<sup>3</sup><https://github.com/peiyunh/tiny>

<sup>4</sup><https://www.ics.uci.edu/~xzhu/face/>

<sup>5</sup><https://github.com/ladrianb/2D-and-3D-face-alignment>

<sup>6</sup><https://github.com/patrikhuber/eos>

<sup>7</sup>[https://github.com/anhtran/3dmm\\_cnn](https://github.com/anhtran/3dmm_cnn)

<sup>8</sup><https://github.com/GilLevi/AgeGenderDeepLearning>

<sup>9</sup><https://github.com/TadasBaltrusaitis/OpenFace>

## XII. FINAL REMARKS

A survey on the state-of-the-art regarding facial image analysis in the wild was presented. The subject has been evolving considerably in the past five years, but the yearning for better results in real-life applications is still high. The performance of current approaches to solve problems due to unconstrained, heterogeneous scenarios is limited by low image quality, occlusions, and varying face poses, which impose a great drawback for automated systems. However, advances have been made in diverse topics related to facial image analysis in the wild. Future work is expected to tackle challenging scenarios where subjects actively try to not be recognized via multiple methods, including blending in a large

<sup>10</sup><https://github.com/zkl20061823/DRML>

<sup>11</sup><https://github.com/kaltwang/latenttrees>

<sup>12</sup>[https://github.com/RWalecki/copula\\_ordinal\\_regression](https://github.com/RWalecki/copula_ordinal_regression)

<sup>13</sup>[http://www.robots.ox.ac.uk/~vgg/software/vgg\\_face/](http://www.robots.ox.ac.uk/~vgg/software/vgg_face/)

<sup>14</sup><https://cmusatyalab.github.io/openface/>

crowd, increasing the search scale, or wearing multiple face accessories, creating occlusions. Facial expression analysis and 3D face recognition, which still heavily rely on controlled acquisitions and dense annotations, should experience a switch to in the wild environments as new, large, and annotated datasets become available.

#### ACKNOWLEDGMENT

The authors would like to thank CAPES and CNPq for supporting this research.

#### REFERENCES

- [1] S. Zafeiriou, G. Trigeorgis, G. Chrysos, J. Deng, and J. Shen, "The menpo facial landmark localisation challenge: A step towards the solution," in *IEEE CVPR Workshops*, July 2017.
- [2] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE TIFS*, 2014.
- [3] D. Wang, C. Otto, and A. K. Jain, "Face search at scale," *IEEE TPAMI*, 2017.
- [4] C. F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *IEEE CVPR*, 2016.
- [5] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," 2016.
- [6] R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa, "An all-in-one convolutional neural network for face analysis," in *FG*. IEEE, 2017.
- [7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE CVPR*, 2001.
- [8] S. Liao, A. K. Jain, and S. Z. Li, "Unconstrained face detection," *TR MSU-CSE*, 2012.
- [9] —, "A fast and accurate unconstrained face detector," *IEEE TPAMI*, 2016.
- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *NIPS*, C. C. et al., Ed., 2015.
- [11] M. Köstinger, P. Wohlhart, P. M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization," in *IEEE ICCV*, 2011.
- [12] H. Jiang and E. Learned-Miller, "Face detection with the faster r-cnn," in *FG*. IEEE, 2017.
- [13] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," 2016.
- [14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *ECCV*. Springer, 2016.
- [15] P. Hu and D. Ramanan, "Finding tiny faces," 2017.
- [16] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, 2014.
- [17] A. Dutta, R. Veldhuis, and L. Spreeuwiers, *A Bayesian model for predicting face recognition performance using image quality*. IEEE, 2014, pp. 1–8.
- [18] L. P. Silva, F. H. B. Zavan, O. R. P. Bellon, and L. Silva, "Follow that nose: tracking faces based on the nose region and image quality feedback," in *Conf. on Graphics, Patterns and Images - W. Face Processing*, 2016.
- [19] F. H. B. Zavan, A. C. P. Nascimento, O. R. P. Bellon, and L. Silva, "Nosepose: a competitive, landmark-free methodology for head pose estimation in the wild," in *Conf. on Graphics, Patterns and Images - W. Face Processing*, 2016.
- [20] J. Shen, S. Zafeiriou, G. G. Chrysos, J. Kossaifi, G. Tzimiropoulos, and M. Pantic, "The first facial landmark tracking in-the-wild challenge: Benchmark and results," in *2015 IEEE ICCV Workshops*, Dec 2015, pp. 1003–1011.
- [21] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Learning face hallucination in the wild," in *AAAI Conf. Artificial Intelligence*, 2015.
- [22] S. Baker and T. Kanade, "Hallucinating faces," *IEEE FG*, 2000.
- [23] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, "A comprehensive survey to face hallucination," *IJCV*, vol. 106, no. 1, pp. 9–30, 2014.
- [24] J. Jiang, R. Hu, Z. Wang, and Z. Han, "Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning," *IEEE TIP*, vol. 23, no. 10, pp. 4220–4231, 2014.
- [25] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "Srlsp: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 27–40, 2017.
- [26] E. Murphy-Chutorian and M. M. Trivedi, *IEEE TPAMI*.
- [27] F. H. de Bittencourt Zavan, A. C. Nascimento, L. P. e Silva, O. R. Bellon, and L. Silva, "3d face alignment in the wild: A landmark-free, nose-based approach," in *ECCV*, 2016.
- [28] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *CVPR*. IEEE, 2012.
- [29] K. Sundararajan and D. L. Woodard, "Head pose estimation in the wild using approximate view manifolds," in *IEEE CVPR Workshops*, 2015, pp. 50–58.
- [30] X. Peng, J. Huang, Q. Hu, S. Zhang, and D. N. Metaxas, "Three-dimensional head pose estimation in-the-wild," in *IEEE FG*, vol. 1. IEEE, 2015, pp. 1–6.
- [31] S. Z. Li and A. Jain, *Encyclopedia of Biometrics*. Springer, 2015.
- [32] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: database and results," *Image and Vision Computing*, 2016.
- [33] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *IEEE CVPR*.
- [34] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *PNAS*, 2014.
- [35] L. Silva, "Rastreamento facial e refinamento de pontos fiduciais 3d baseado na região do nariz em ambientes não controlados," in *Dissertação de Mestrado, Universidade Federal do Paraná - UFPR*, 2017.
- [36] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," *IJCV*, 2014.
- [37] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in *IEEE CVPR*, 2013.
- [38] S. Zhu, C. Li, C. Change Loy, and X. Tang, "Face alignment by coarse-to-fine shape searching," in *IEEE CVPR*, 2015.
- [39] J. Yan, Z. Lei, D. Yi, and S. Li, "Learn to combine multiple hypotheses for accurate face alignment," in *IEEE ICCV*, 2013.
- [40] H. Yang, W. Mou, Y. Zhang, I. Patras, H. Gunes, and P. Robinson, "Face alignment assisted by head pose estimation," in *BMVC*, 2015.
- [41] G. Trigeorgis, P. Snape, M. A. Nicolaou, E. Antonakos, and S. Zafeiriou, "Mnemonic descent method: A recurrent process applied for end-to-end face alignment," in *IEEE CVPR*, 2016.
- [42] L. A. Jeni, S. Tulyakov, L. Yin, N. Sebe, and J. F. Cohn, "The first 3d face alignment in the wild (3dfaw) challenge," in *ECCV*, 2016.
- [43] R. Zhao, Y. Wang, C. F. Benitez-Quiroz, Y. Liu, and A. M. Martinez, "Fast and precise face alignment and 3d shape reconstruction from a single 2d image," in *ECCV*, 2016.
- [44] A. Bulat and G. Tzimiropoulos, "Two-stage convolutional part heatmap regression for the 1st 3d face alignment in the wild (3dfaw) challenge," in *ECCV*, 2016.
- [45] A. T. Tran, T. Hassner, I. Masi, and G. Medioni, "Regressing robust and discriminative 3d morphable models with a very deep neural network," *IEEE CVPR*, 2017.
- [46] P. Huber, Z.-H. Feng, W. Christmas, J. Kittler, and M. Ratsch, "Fitting 3d morphable face models using local features," in *ICIP*. IEEE, 2015, pp. 1195–1199.
- [47] V. S. Ramachandran, "Perception of shape from shading," *Nature*, vol. 331, no. 6152, pp. 163–166, 1988.
- [48] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3d face model for pose and illumination invariant face recognition," in *IEEE Conf. on Advanced Video and Signal Based Surveillance*, 2009.
- [49] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3d morphable model learnt from 10,000 faces," in *IEEE CVPR*, 2016.
- [50] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE TPAMI*, 2003.
- [51] S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3d morphable model using linear shape and texture error functions," *ECCV*, 2006.
- [52] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *IEEE TPAMI*, 2011.

- [53] S. Liang, L. G. Shapiro, and I. Kemelmacher-Shlizerman, "Head reconstruction from internet photos," in *ECCV*, 2016.
- [54] M. Pietraschke and V. Blanz, "Automated 3d face reconstruction from multiple images using quality measures," in *IEEE CVPR*, 2016.
- [55] S. Escalera, J. Fabian, P. Pardo, X. Baró, J. González, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon, "Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *IEEE ICCVW*, 2015, pp. 243–251.
- [56] E. Eiding, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE TIFS*, no. 12, pp. 2170–2179, 2014.
- [57] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *CVPR*. IEEE, 2015.
- [58] J. Mansanet, A. Albiol, and R. Paredes, "Local deep neural networks for gender recognition," *Pattern Recognition Letters*, vol. 70, pp. 80–86, 2016.
- [59] Z. Lou, F. Alnajar, J. Alvarez, N. Hu, and T. Gevers, "Expression-invariant age estimation using structured learning," *IEEE TPAMI*, vol. PP, no. 99, pp. 1–1, 2017.
- [60] P. Ekman, W. Friesen, and J. Hager, *Facial Action Coding System (FACS): Manual*. A Human Face, 2002.
- [61] W. V. F. Paul Ekman, *Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions*.
- [62] J. A. Russel, "A circumplex model of affect," *Journal of Personality and Social Psychology*, 1980.
- [63] S. Du and A. M. Martinez, "Compound facial expressions of emotion: from basic research to clinical applications," *Dialogues in Clinical Neuroscience*, 2015.
- [64] T. Kanade, Y. Tian, and J. F. Cohn, "Comprehensive database for facial expression analysis," in *IEEE FG*, 2000.
- [65] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database," in *IEEE FG*, 2008.
- [66] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn, "Disfa: A spontaneous facial action intensity database," *IEEE Trans. on Affective Computing*, 2013.
- [67] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database," *Image and Vision Computing*, 2014.
- [68] Z. Zhang, J. M. Girard, Y. Wu, X. Zhang, P. Liu, U. Ciftci, S. Canavan, M. Reale, A. Horowitz, H. Yang, J. F. Cohn, Q. Ji, and L. Yin, "Multimodal spontaneous emotion corpus for human behavior analysis," in *IEEE CVPR*, 2016.
- [69] T. Baltrušaitis, P. Robinson, L.-P. Morency *et al.*, "Openface: an open source facial behavior analysis toolkit," in *IEEE WACV*, 2016.
- [70] J. C. Batista, V. Albiero, O. R. P. Bellon, and L. Silva, "Aumpnet: Simultaneous action units detection and intensity estimation on multi-pose facial images using a single convolutional neural network," in *IEEE FG*, 2017.
- [71] K. Zhao, W.-S. Chu, and H. Zhang, "Deep region and multi-label learning for facial action unit detection," in *IEEE CVPR*, 2016.
- [72] R. Walecki, O. Rudovic, V. Pavlovic, and M. Pantic, "Copula ordinal regression for joint estimation of facial action unit intensity," in *IEEE CVPR*, 2016.
- [73] J. M. Girard, J. F. Cohn, and F. D. la Torre, "Estimating smile intensity: A better way," *Pattern Recognition Letters*, 2015.
- [74] S. Kaltwang, S. Todorovic, and M. Pantic, "Latent trees for estimating intensity of facial action units," in *IEEE CVPR*, 2015.
- [75] B. Martinez, M. F. Valstar, B. Jiang, and M. Pantic, "Automatic analysis of facial actions: A survey," *IEEE TAFCC*, 2017.
- [76] C. F. Benitez-Quiroz, R. Srinivasan, Q. Feng, Y. Wang, and A. M. Martinez, "Emotionet challenge: Recognition of facial expressions of emotion in the wild," *arXiv preprint arXiv:1703.01210*, 2017.
- [77] M. Valstar, E. S. Lozano, J. F. Cohn, L. A. Jeni, J. M. Girard, L. Yin, Z. Zhang, and M. Pantic, "Fera 2017 - addressing head pose in the third facial expression recognition and analysis challenge," in *IEEE FG*, 2017.
- [78] Y. Zhou, J. Pi, and B. E. Shi, "Pose-independent facial action unit intensity regression based on multi-task deep transfer learning," in *FG*. IEEE, 2017.
- [79] C. Tang, W. Zheng, J. Yan, Q. Li, Y. Li, T. Zhang, and Z. Cui, "View-independent facial action unit detection," in *FG*. IEEE, 2017.
- [80] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *BMVC*, 2015, pp. 41.1–41.12.
- [81] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, 2016.
- [82] I. Masi, T. A. Tuán, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in *ECCV*, 2016, pp. 579–596.
- [83] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2010-009, 2010.
- [84] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *IEEE CVPR*, 2016.
- [85] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *IMAVIS*, vol. 28, no. 5, pp. 807–813, 2010.
- [86] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," *BIOID*, pp. 47–56, 2008.
- [87] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks)," in *IEEE ICCV*, 2017.
- [88] S. Escalera, M. Torres Torres, B. Martinez, X. Baro, H. Jair Escalante, I. Guyon, G. Tzimiropoulos, C. Corneou, M. Oliu, M. Ali Bagheri, and M. Valstar, "Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016," in *IEEE CVPR Workshops*, June 2016.
- [89] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *IEEE ICCV*, 2015.
- [90] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE TAFCC*, 2017.
- [91] A. Mollahosseini, B. Hasani, M. J. Salvador, H. Abdollahi, D. Chan, and M. H. Mahoor, "Facial expression recognition from world wild web," in *IEEE CVPR Workshops*, 2016.
- [92] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," *Neural Networks*, 2015.
- [93] A. Dhall, R. Goecke, J. Joshi, and T. Gedeon, "Emotion recognition in the wild challenge 2016," in *ACM ICMI*, 2016.
- [94] M. Valstar, J. Gratch, B. Schuller, F. Ringeval, R. Cowie, and M. Pantic, "Summary for avec 2016: Depression, mood, and emotion recognition workshop and challenge," in *ACM Multimedia Conference*, 2016.
- [95] S. Zafeiriou, A. Papaioannou, I. Kotsia, M. Nicolaou, and G. Zhao, "Facial affect in-the-wild: A survey and a new database," in *IEEE CVPRW*, 2016.
- [96] D. McDuff, R. el Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. Picard, "Affectiva-mit facial expression dataset (am-fed): Naturalistic and spontaneous facial expressions collected in-the-wild," in *IEEE CVPR*, 2013.
- [97] M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, and K. Scherer, "The first facial expression recognition and analysis challenge," in *IEEE FG*, 2011.
- [98] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn, "Fera 2015 - second facial expression recognition and analysis challenge," in *IEEE FG*, 2015.
- [99] J. Beveridge, P. Phillips, D. Bolme, B. Draper, G. Givens, Y. M. Lui, M. Teli, H. Zhang, W. Scruggs, K. Bowyer, P. Flynn, and S. Cheng, "The challenge of face recognition from digital point-and-shoot cameras," in *IEEE BTAS*, 2013.
- [100] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, Tech. Rep., 2007.
- [101] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *IEEE CVPR*, 2011, pp. 529–534.
- [102] S. Suwajanakorn, I. Kemelmacher-Shlizerman, and S. M. Seitz, "Total moving face reconstruction," in *ECCV*. Springer, 2014, pp. 796–812.