What is the relationship between face alignment and facial expression recognition?

Romain Belmonte [1], Benjamin Allaert [1], Pierre Tirilly [1], Ioan Marius Bilasco [1], Chaabane Djeraba [1] and Nicu Sebe [2]

[1] Centre de Recherche en Informatique Signal et Automatique de Lille, Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL -, F-59000 Lille, France.

[2] Department of Information Engineering and Computer Science (DISI), University of Trento, Italy.

Version May 28, 2019 submitted to Journal Not Specified

Abstract: Face expression recognition is still a complex task, particularly due to the presence of head pose variations. Although face alignment approaches are becoming increasingly accurate for characterizing facial regions, it is important to consider the impact of these approaches when they are used for other related tasks such as head pose registration or facial expression recognition. In this paper, we compare the performance of recent face alignment approaches to highlight the most appropriate techniques for preserving facial geometry when correcting the head pose variation. Also, we highlight the most suitable techniques that locate facial landmarks in the presence of head pose variations and facial expressions.

Keywords: Face alignment, head pose registration, facial expressions.

1. Introduction

Despite continuous progress in face analysis, numerous studies still focus on near-frontal faces. As a result, most approaches have some difficulties performing well under conditions where faces appear under a wide range of poses (e.g., in video surveillance systems). Initially focused on constrained pose conditions, new approaches tend to consider the whole range of facial poses in situations of natural interactions. This constitutes a new challenging trend in face analysis [1,2].

Face alignment approaches have proven their effectiveness in detecting the face and locating the different facial elements (eyes, nose, and mouth). In addition to it, these techniques can be subsequently used to deal with head pose variations [3] and support various facial analysis tasks, such as expression recognition [4]. Based on the distribution of landmarks, the face can be registered in order to guarantee stable locations for the major facial components across different face images, and minimize the variations in the scales, rotations, and position of the faces. Once this spatial invariance has been achieved, it is easier to extract robust descriptors characterizing the face.

Over the years, many competitions have invited researchers to develop more robust algorithms for facial landmark detection by addressing a wide range of situations (head poses, facial expressions, illumination, etc.). Dealing with these situations makes the identification of facial landmarks more robust. However, there is no guarantee that an approach providing more accurate detections always leads to better final performances when used for subsequent tasks such as facial expression recognition, as illustrated in Figure 1.

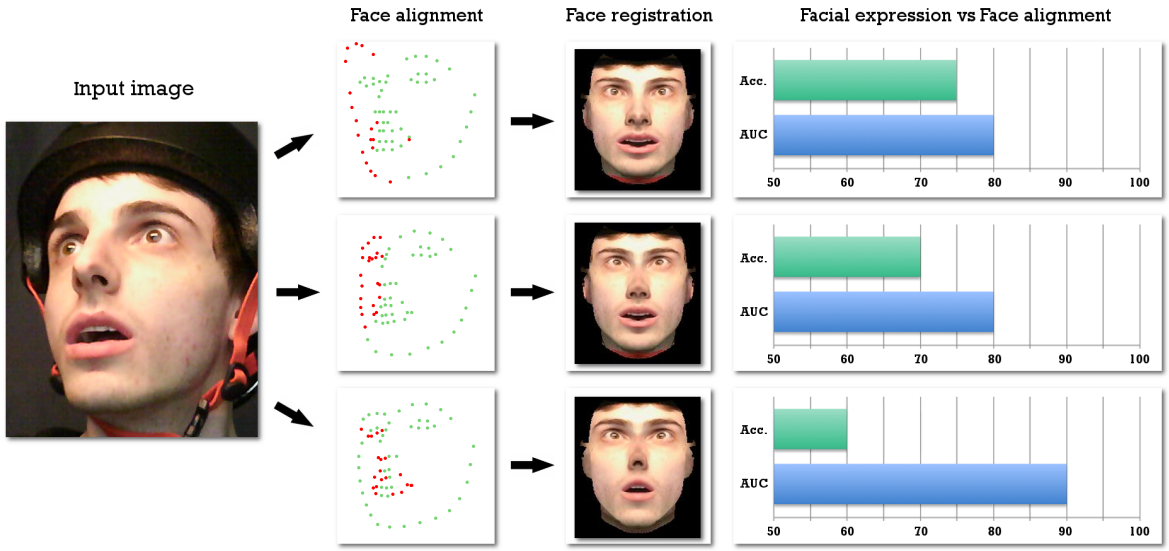

Figure 1. Comparison of the ability of different face alignment approaches to maintain facial geometry when used to register the face. On the right side of the image we illustrate the registered face using the landmarks provided by each face alignment solution considered. The graphs present a comparison of face alignment performance (measured as Area Under Curved - AUC) and facial expression recognition performance (measured as accuracy rate - Acc.). A higher average AUC does not guarantee that the resulting alignment is better suited to recognize facial expressions, mainly because some landmarks have a greater impact on face registration. More significant geometric deformations can be induced if the alignment fails on some of the important landmarks (red dots).

This study is the first attempt to address the questions "What effect does face alignment have on facial expression recognition?" and "What effect do facial expressions have on face alignment?", which are of increasing interest in computer vision community [1]. We contribute to answering this question by comparing some of the recent face alignment approaches to identify whether the performance criteria used in current competitions are relevant when considering face alignment as a tool for specific tasks such as facial expression recognition.

Section 2 highlights the main objectives of our paper and presents the evolution of recent methods for face alignment and facial expression recognition. Section 3 introduces the dataset used to compare the different face alignment approaches and the performance criteria applied in the experiments. Section 4 evaluates the ability of the selected face alignment approaches to locate facial landmarks in the presence of varying head poses and facial expressions. Section 5 assesses the ability of face alignment approaches to maintain facial geometry when used for face expression recognition. The results and perspectives are discussed in Section 6.

2. Background and scope

This section highlights the main objectives of the paper and presents a brief overview of existing methods for face alignment and facial expression recognition.

Fig. 2 shows the typical process for facial expression recognition. Facial components are usually detected by face alignment approaches, represented by a distribution of key points (typically, 68 facial landmarks). Based on this distribution, the face can be registered in order to guarantee stable locations for the major facial components across different face images. The face is registered to minimize the variations in its scale, rotation, and position over images. Once the face is registered, features characterizing expression-related facial deformations are extracted and facial expression recognition is performed, typically using a supervised classifier.

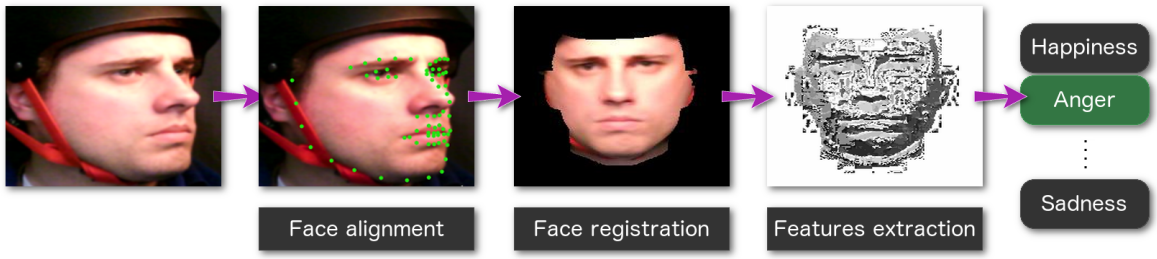

Figure 2. Typical process for facial expression recognition in a natural interaction situation.

2.1. Face alignment

The majority of face alignment approaches are based on cascaded regression [5]. It is a coarse-to-fine strategy that consists in progressively updating the positions of landmarks through regression functions learned directly from features representing the appearance of the face. Today, feature extraction and regression are trained jointly using deep neural networks. Two main architectures can be distinguished: a) networks that directly regress landmark coordinates using a fully connected layer, and b) fully convolutional networks (i.e., without any fully connected layer) that regress heatmaps, one for each landmark. The latter has become popular, especially through hourglass-like architectures, which stack encoder-decoder networks with intermediate supervision to better capture spatial relationships [6]. Landmark heatmaps can also be used to transfer information between stages during cascading regression using coordinate regression [7]. Face alignment does not necessarily have to be treated independently and can be learned together with correlated facial attributes with multi-task networks [8]; it helps achieving individual performance gains on each task. While most authors focus on the variance of faces, the intrinsic variance of image styles can also be handled to improve performance using style-aggregated networks [9].

Recent work has shown that temporal coherence can be used to cope with facial and environmental variability under uncontrolled conditions. The most recent methods generally combine convolutional neural networks (CNNs) and recurrent neural networks (RNNs) while decoupling the processing of spatial and temporal information to better leverage their complementarity [10–12]. An RNN estimates and tracks jointly visual

features over time. This late temporal connectivity helps stabilize predictions and handle head pose variations [12,13]. An unsupervised approach based on the coherency of optical flow can encourage temporal consistency in image-based detectors, which can reduce jittering in videos [14]. The statistics of different kinds of movements can be learned using a stabilization model coupled with a loss function including a regularization term and a smoothing term to address time delays and smoothness issues [15]. To go further, local motion can be included using early temporal connectivity based on 3D convolutions [16]. By improving the temporal connectivity, more accurate predictions can be obtained, especially during expression variations.

Another trend is the use of depth information to improve the accuracy of landmarks [6,17–19]. The vast majority of methods consider the face as a 2D object; it is not so surprising to see that out-of-plane rotations are an issue for these methods. To overcome it, 3D landmarks can be computed from 2D ones [6]. For instance, a 3D Morphable Model (3DMM) can be fit to 2D facial images [18,19]. More recently, 3D landmarks can also be directly estimated from 2D facial images [6,17]. These methods are generally based on a cascade of CNNs. Depth and temporal informations are not mutually exclusive and methods leveraging both of them could help to improve the robustness of facial alignment.

2.2. Facial expression recognition

Skin deformations induced by face muscles characterize facial expressions. In facial deformation analysis, several types of techniques exist to encode these changes. They can be based on appearance features, geometry features, or both.

Several appearance features have been proposed such as local binary patterns (LBP) [20]. They provide good results in the analysis of macro facial deformations. CNN-based approaches [21] perform well too, when they learn spatial features from apex frames (i.e. the frames of a video that depict the expressions at their highest intensity). By relying on spatial features only, LBP and static CNN approaches do not utilize the dynamics of facial expressions to recognize them, which can limit their performances at non-apex frames or in the presence of subtle expressions.

Psychological experiments by Bassili [22] showed that facial expressions are recognized more accurately in sequences of images. Therefore, a dynamic extension of LBP, called local binary pattern on three orthogonal plans (LBP-TOP), is proposed in [23]. In the same line of work, and considering the latest developments in dynamic texture modeling, optical flow has regained interest from the community, becoming one of the most widely used solutions [4]. Although temporal approaches tend to provide good performance, they are very sensitive to the noise caused by facial deformations or head movements.

All these approaches have proven their effectiveness in characterizing facial expressions on static and frontal faces. However, facial expression analysis in natural interaction situations (i.e. unconstrained pose settings) is a complex issue. It requires algorithms to be invariant to head pose variations (involving in-plane and out-of-plane rotations) and large head displacements (involving large in-plane translations). To do so, face alignment approaches are used to bring the face into an ideal setting (typically, a frontal pose). Eye registration is

the most popular strategy in near frontal-view databases. The limit of this approach is that eyes must be detected well in the first place. Extensions considering more landmarks are supposed to provide a greater stability when individual landmarks are poorly detected. Methods based on 2D features [24] are suitable for the analysis of near-frontal facial expressions in the presence of limited head motions. But, they do not cope well with occlusions and out-of-plane rotations. Recent approaches propose a robust landmark-based registration using 3D models [3] to generate natural face images in a frontal pose. Compared to 2D approaches, 3D approaches reduce the deformations of the face when facial expressions occur.

2.3. Scope of the paper

In this paper, we first evaluate the robustness of recent face alignment approaches to head pose variations and facial expressions. Then, we investigate the impact of face alignment on expression recognition.

During natural interactions, a misalignment of the facial landmarks often occurs. This is primarily caused by variations in head pose and by facial deformations induced by expressions. Indeed, in the presence of certain head poses, some regions of the face tend to disappear, increasing the difficulty of facial landmark detection. As for expressions, some movements induce complex deformations (typically, around important facial elements such as lips), which also impede landmark detection.

Over the years, many datasets have helped researchers increase the robustness of face alignment approaches. Although these datasets can feature a large range of variations (head poses, facial expressions, illumination, etc.), they do not allow the accurate identification and measurement of the weaknesses of face alignment approaches in relation to a given factor, such as head pose variations or facial expressions. They lack suitable annotations and they do not contain aligned data captured in the absence of the variations, which is a required setting for assessing the impact of a single factor. With the emergence of new datasets, such as SNaP-2DFe [25], that provide synchronized and accurate labels of facial landmarks in both the presence and the absence of specific factors (e.g., head pose variations, facial expressions), it is possible to measure the robustness of face alignment approaches to these factors.

Through our evaluation, and for the first time, we discuss two aspects:

1. the quality of facial landmark detection in the simultaneous presence of head pose variations and facial expressions;
2. the impact of face landmark detection on a subsequent expression recognition process.

3. Experimental conditions

After this brief review of the major approaches to face alignment and facial expression recognition, we now proceed to carrying out a comprehensive comparison of landmark localization and expression recognition performances. In this section, we first introduce the dataset, the face alignment approaches, and the expression recognition methods that we selected for our experiments, then, the evaluation criteria that we use.

3.1. Selected dataset

In this paper, we use the SNaP-2DFe dataset. Unlike other facial expression datasets, SNaP-2DFe offers the possibility to analyze the impact of facial expressions on face alignment, and vice-versa. Indeed, the data in SNaP-2DFe has been collected simultaneously under constrained and unconstrained head poses, as illustrated in Fig. 3. It makes it possible to highlight the facial deformations induced by the face registration step, which depends on the facial landmarks provided by the face alignment approaches.

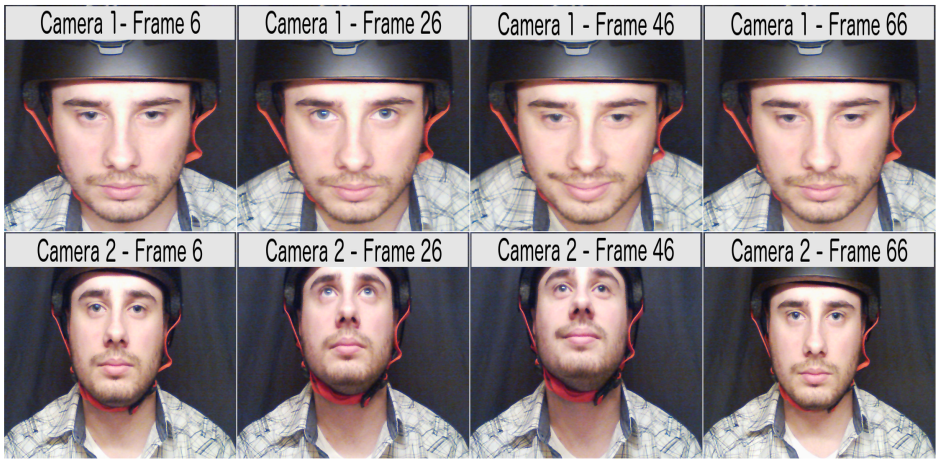

Figure 3. Sample images of facial expressions recorded under pitch movements from the SNaP-2DFe dataset (row 1: helmet camera, expression movement only; row 2: static camera, both expression and head movements).

The SNaP-2DFe dataset contains more than 93,240 images from 1260 videos of 15 subjects eliciting various facial expressions. These videos contain synchronized image sequences of faces in frontal and non-frontal situations. For each subject, six head pose variations (Static – no head movement, translation along the X axis (Tx), Roll, Yaw, Pitch, Roll, and diagonal (Diag) – from the upper-left corner to the lower-right corner) combined with seven expressions (Neutral, Anger, Disgust, Fear, Happiness, Sadness, and Surprise) were recorded by two cameras, resulting in a total of 630 constrained recordings (i.e., recordings without head movements) and 630 unconstrained recordings (i.e., recordings with head movements). SNaP-2DFe provides temporal annotations of the temporal patterns of expression activation (neutral-onset-apex-offset-neutral). Sixty-eight facial landmark locations have been initially extracted using the method of Kazemi and Sullivan [26]. All frames were then individually inspected and, when needed, re-annotated in order to compensate for landmark estimation errors.

3.2. Selection of face alignment approaches

Given the large number of face alignment approaches in the literature, we have selected only a representative subset of recent approaches. We focus on approaches based on deep learning as they currently constitute the dominant trend. Among them, we selected state-of-the-art models for each of the categories that we highlighted in Section 2.1:

- coordinate regression models: DAN [7]);
- heatmap regression models: HG [6] and SAN [9];

- multi-task models: TCDCN [8];
- dynamic models: SBR [14] and FHR [15].

These approaches mostly use image collections as training data (see Table 1). The sizes of the datasets are variable. So, generally, authors combine several datasets to make a larger dataset, which is necessary to deal with the large range of possible head pose variations and facial expressions. It should be noted that the majority of approaches use the 300W, HELEN, and AFLW datasets. Others such as HG, TCDCN, SBR, and FHR use additional training sets to improve the robustness of their model.

Table 1. Datasets used to train the different approaches selected for the evaluations.

Datasets			Face alignment approaches					
Name	Type	Content	HG	TCDCN	DAN	FHR	SBR	SAN
300W [27]	Static	600 img	✓	✓	✓	✓	✓	✓
HELEN [28]	Static	2,330 img	-	✓	✓	✓	✓	✓
AFLW [29]	Static	25,000 img	✓	✓	✓	✓	✓	✓
COFW [30]	Static	1007 img	-	✓	-	-	-	-
MAFL [8]	Static	20,000 img	-	✓	-	-	-	-
300W-LP [31]	Static	61,225 img	✓	-	-	-	-	-
Menpo [32]	Static	14,854 img	✓	-	✓	-	-	-
300VW [33]	Temporal	114 seq / 218,595 img	✓	-	-	✓	✓	-

It is important to note that, although the 300VW dataset contains temporal data (video sequences), most approaches that use it do not use this information.Furthermore, few datasets provide 3D landmarks annotations. So, we focus our study on comparing approaches based on static 2D approaches.

In the following evaluation, we use the code and pre-trained models provided by the authors. We do not perform any fine-tuning on SNaP-2DFe. We do so because we aim to assess the generalization capability of the landmark detection models and their fitness for subsequent tasks (here, facial expression recognition); we believe one should not have to re-train alignment models specifically for their application.

3.3. Selection of facial expression recognition approaches

Based on the landmarks provided by face alignment approaches, one face registration technique is used on the subset of SNaP-2DFe recorded by the static camera in order to correct head pose variations and obtain frontal faces. Among the different face registration approaches used in the literature to deal with head pose, we have applied the recent 3D approach proposed by Hassner et al. [3]. This approach has the advantage to preserve facial expressions.

We select two typical features for facial expression recognition: one based on facial appearance – LBP [20] – and one based on facial motion – LMP [4]. We do not include models based on deep learning, for two reasons:

- the lack of training data;
- the fact that these models are usually applied to apex frames only, and offer no guarantee when applied to whole image sequences, from the onset to the offset of the expression.

3.4. Performance criteria

The mean Euclidean distance e between the predicted landmarks and the ground truth normalized by the diagonal of the ground truth bounding box is used as an evaluation metric for its robustness to pose variations [2]. The error e_n for the $n-th$ image is expressed as:

$$e_n = \frac{1}{L}\sum_{i=1}^{L}\frac{||p_i - g_i||_2}{D} \tag{1}$$

where L is the number of landmarks, p_i is the coordinates of the i-th predicted landmark, g_i is the coordinates of the corresponding ground truth landmark, and D is the diagonal of the ground truth bounding box ($D = \text{round}(\sqrt{w^2 + \text{h}^2})$, with w and h the width and height of the ground truth bounding box, respectively). From this metric, we compute the area under the curve (AUC) and the failure rate (FR) with a 0.04 threshold. Above this threshold, we consider a prediction as a failure, since a facial component can be completely mismatched. The AUC and FR are expressed as:

$$\text{AUC}_\alpha = \int_0^\alpha f(e)de \tag{2}$$

$$\text{FR}_\alpha = 1 - f(\alpha). \tag{3}$$

where f is the cumulative error distribution (CED) function and α the threshold.

Facial expression recognition is performed by training a SVM classifier with an RBF kernel and applying a 10-fold cross validation protocol. For each evaluation, we report the average cross-validation accuracy.

4. Effectiveness of face alignment

In this section, we investigate the robustness of face alignment in the presence of head pose variations and facial expressions. First, the analysis is focused on the complete distribution of facial landmarks in order to identify which conditions challenge the most the approaches studied. Second, an analysis is carried out on each facial landmark individually, in order to identify more precisely the facial regions that are more difficult to characterize due to facial expressions and head pose variations.

4.1. Overall performance analysis

In this experiment, we examine which movements, the ones induced by pose variations or the ones produced by facial expressions, have the largest impact on face alignment. Table 2 presents the performance of the face alignment approaches in the presence of head pose variations only (i.e., Neutral expression / six head movements), facial expressions only (i.e., seven expressions / Static frontal pose) and head pose variations combined with facial expressions (i.e., seven expressions / six head movements). For each face alignment approach, the AUC and the FR are calculated from the first image to the last.

Table 2 shows that the AUC is lower and the FR is higher in the presence of head pose variations than in the presence of facial expressions, for all face alignment approaches. This result outlines that head pose variations are

Table 2. AUC/FR with or without head pose variations and facial expressions on SNaP-2DFe.

	Pose variations only	Expressions only	Pose variations & expressions	Average
HG	54.30 / 0.42	55.35 / 0.05	52.94 / 0.58	54.19 / 0.35
TCDCN	54.86 / 4.49	59.73 / 0.00	52.97 / 5.01	55.85 / 3.16
DAN	68.73 / 8.44	71.77 / 6.41	67.88 / 8.34	69.46 / 7.73
FHR	69.90 / 0.17	71.84 / 0.00	68.52 / 0.64	70.08 / 0.27
SBR	70.44 / 1.02	73.76 / 0.57	68.86 / 2.02	71.02 / 1.20
SAN	70.97 / 0.21	73.80 / 0.00	70.27 / 0.44	71.68 / 0.21
Average	64.87 / 2.46	67.71 / 1.17	63.57 / 2.84	65.38 / 2.15

more challenging than facial expressions for face alignment. When both challenges are present, the performance tends to decrease further.

Table 3 provides a more detailed view of the results with both challenges (head pose variations and expressions) by dividing them according to the activation patterns of the facial expression (four periods: neutral to onset, onset to apex, apex to offset, offset to neutral).

Table 3. AUC/FR by activation pattern on SNaP-2DFe dataset in the presence of facial expression and head pose variations.

	Neutral / Onset	Onset / Apex	Apex / Offset	Offset / Neutral	All
HG	55.21 / 0.06	49.30 / 1.73	48.58 / 1.56	51.74 / 0.65	52.94 / 0.58
TCDCN	55.55 / 3.17	44.70 / 11.77	45.99 / 9.64	51.84 / 4.92	52.97 / 5.01
DAN	69.87 / 7.91	62.35 / 9.95	64.07 / 8.99	67.51 / 8.34	67.88 / 8.34
FHR	71.57 / 0.04	64.14 / 2.37	63.24 / 1.67	67.15 / 0.62	68.52 / 0.64
SBR	71.74 / 1.04	62.22 / 5.12	63.75 / 3.42	67.35 / 2.83	68.86 / 2.02
SAN	72.79 / 0.00	65.65 / 1.22	66.41 / 1.08	68.86 / 0.56	70.27 / 0.44
Average	66.12 / 2.04	58.06 / 5.36	58.67 / 4.39	62.41 / 2.99	63.57 / 2.84

In Table 3, the accuracies of all face alignment methods decrease the most for images adjacent to the apex state. It corresponds to the moment when the expression and most head pose variations are at their highest intensity. As soon as the subject gets closer to a neutral expression and a frontal pose, the accuracy of face alignment improves. This result confirm that expressions with head pose variations remain a major difficulty for face alignment, and shows that not only the presence of an expression, but also its intensity, impact the alignment.

4.2. Robustness to head pose variations

In this experiment, we focus on head pose variations only, by investigating which types of head poses are the most challenging. To do so, we run experiments only on sequences where the face has a neutral expression from the onset to the offset. In Table 4, the results are split by type of head pose variations. The reported figures (ΔAUC and ΔFR) are the differences in performance from static Neutral faces to Neutral faces in the presence of head pose variations. Negative values (resp. positive values) for ΔAUC indicate a decrease (resp. increase) in performance due to the head pose. Similarly, positive values (resp. negatives values) for ΔFR indicate a decrease (resp. increase) in performance due to the head pose.

The results in Table 4 show that some head pose variations impede face alignment more than the others. Diag and Pitch lead to severe drops of the AUC and Diag increases the FR considerably, suggesting that these

Table 4. ΔAUC / ΔFR for each type of head pose variations on SNaP-2DFe.

	Tx	Roll	Yaw	Pitch	Diag	Average
HG	-0.89 / 0.0	+0.67 / 0.0	-4.17 / 0.0	-1.19 / +0.67	-10.18 / +3.67	-3.15 / +0.86
TCDCN	-0.36 / 0.0	-1.36 / 0.0	-21.33 / +24.0	-9.39 / 0.0	-30.65 / +33.67	-12.61 / +11.53
DAN	-4.0 / +2.0	-13.15 / +15.0	-1.76 / -1.0	-8.8 / +0.66	-15.79 / +4.33	-8.7 / +4.19
FHR	-0.97 / 0.0	-1.69 / 0.0	-0.66 / 0.0	-10.08 / 0.0	-7.5 / +2.0	-4.18 / +0.4
SBR	-4.21 / -0.33	-2.82 / -1.0	-7.15 / -0.33	-12.36 / +0.0	-16.71 / +1.67	-8.65 / 0.0
SAN	-0.75 / 0.0	-5.22 / 0.0	-2.36 / 0.0	-8.02 / 0.0	-11.08 / 0.0	-5.48 / 0.0
Average	-1.86 / +0.28	-3.93 / +2.33	-6.24 / +3.78	-8.31 / +0.22	-15.32 / +7.56	-7.13 / +2.83

pose variations are the most challenging. They involve out-of-plane rotations, which have a significant impact on the appearance of the face. The difference in FR (stable for Pitch, increased for Diag) shows that the combination of several out-of-planes rotations in the Diag movement makes landmark detection more difficult, resulting in more totally mismatched detections. Yaw, Roll, and Tx also decrease the AUC, but with more stable values of the FR,[1] showing that face alignment approaches can manage these variations better. Despite the decrease in AUC, the stable FR shows that the errors generated are fairly small and do not result in landmark detection failures.

4.3. Robustness to facial expressions

In this experiment, we focus on facial expressions only, by investigating which categories of facial expressions are the most challenging. To do so, the results are computed only on sequences where the head is frontal and static from the onset to the offset of the facial expression. Table 5 presents the results split by facial expression category. The figures correspond to differences in performance from static neutral faces to faces displaying facial expressions. Negative values (resp. positive values) for ΔAUC indicate a decrease (resp. increase) in performance due to the head pose. Similarly, positive values (resp. negatives values) for ΔFR indicate a decrease (resp. increase) in performance due to the head pose.

Table 5. ΔAUC / ΔFR for each category of facial expressions on SNaP-2DFe.

	Happiness	Anger	Disgust	Fear	Surprise	Sadness	Average
HG	-4.85 / +0.36	-4.41 / 0.0	-8.68 / 0.0	-1.75 / 0.0	-1.35 / 0.0	-3.39 / 0.0	-4.07 / 0.06
TCDCN	-3.12 / 0.0	+1.51 / 0.0	-8.1 / 0.0	-0.02 / 0.0	+0.55 / 0.0	-4.57 / 0.0	-2.29 / 0.0
DAN	+0.48 / -4.30	-0.86 / -0.22	-7.04 / +0.74	-1.29 / -2.16	-1.21 / -0.12	-3.11 / -2.67	-2.17 / -1.45
FHR	-1.44 / 0.0	-2.28 / 0.0	-6.62 / 0.0	-0.96 / 0.0	+0.93 / 0.0	-4.09 / 0.0	-2.41 / 0.0
SBR	+0.04 / -0.64	-0.37 / -1.00	-7.27 / -0.63	-3.49 / +2.01	-0.37 / -1.00	-4.03 / -0.38	-2.58 / -0.27
SAN	+0.34 / 0.0	+1.0 / 0.0	-7.41 / 0.0	-1.52 / 0.0	-1.61 / 0.0	-4.12 / 0.0	-2.22 / 0.0
Average	-1.43 / -0.76	-0.90 / -0.20	-7.52 / +0.02	-1.51 / -0.03	-0.51 / -0.19	-3.89 / -0.51	-2.62 / -0.28

Based on the average results, the three best performances in terms of AUC for face alignment in the presence of facial expressions are DAN, TCDCN and SAN. The results in Table 5 also show that, overall, the performances decrease when facial expressions occur. Disgust and Sadness lead to the most significant drops in the AUC. These expressions involve more complex and more heterogeneous mouth motions and activation sequences with

[1] The average FR for Yaw is due to a single outlier (TCDCN); FR values remain stable for all other methods (with a slight improvement for SBR).

significant changes in appearance, which may explain why face alignment approaches have more difficulties in handling them. The decrease in the AUC is smaller for Happiness, Anger, Fear, and Surprise, which seems to be handled better.

4.4. Analysis of the landmarks

To better identify the strengths and weaknesses of each facial alignment approach, we perform a more detailed analysis at the level of each landmark. Figure 4 provides heatmaps corresponding to the error level per landmark of the different alignment approaches, for each head pose variation (Figure 4-A), for each expression (Figure 4-B), and for the combination of the two (Figure 4-C). For each landmark and for each frame we compute the absolute distance between the estimated location and the ground truth. Distances superior to a threshold corresponding to 0.04 of the face diagonal are associated to a 100% error level. A perfect match results in a 0% error level. Error rates are averaged over all frames for each landmark. With the proposed metric, we quantify errors for each landmark regardless of the global face alignment failure or success with regard to the AUC and FR metrics. The equation 4 illustrates the computation of the error level observed while detecting the i-th landmark over F frames with a threshold error corresponding to a 0.04 factor applied to the face diagonal (D). For instance, the images in column Roll of the Figure 4-A are computed over all Roll sequences where the subjects are expressing Neutral expressions. The images in column Disgusts of the Figure 4-B are computed over all sequences where the head is still (Static) and the subjects are expressing disgusts expressions. The images, in Figure 4-C are computed over all sequences, regardless of the head movements or the expressions.

$$\text{error level}_i(x) = \frac{1}{F}\sum \min(1, \frac{||p_i - g_i||_2}{D * 0.04}) \tag{4}$$

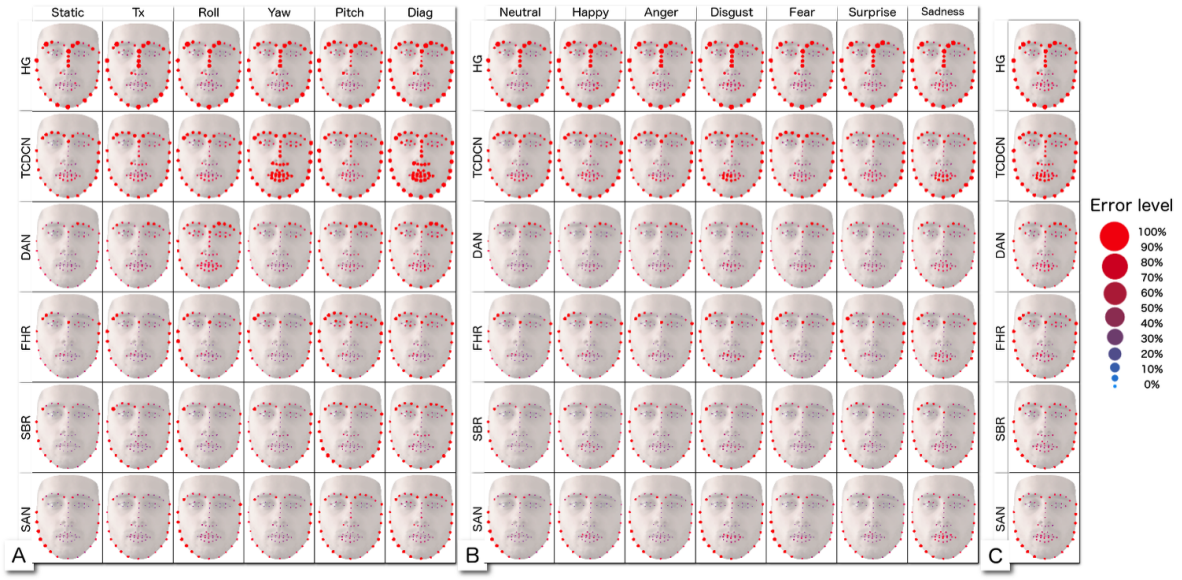

Figure 4. Heatmaps of landmark detection error (A: per head pose, B: per facial expression, C: overall).

Concerning the presence of head pose variations (Figure 4-A), in-plane variations (Static, Tx, and Roll) do not strongly challenge facial alignment approaches, with the exception of the HG and TCDCN models in which landmarks around the eyebrows, the nose, and the facial contours have a relatively high error level. It is important to note that DAN is highly challenged by Roll movements. For out-of-plane (Yaw, Pitch, and Diag) variations, the results are more mixed. HG and TCDCN are even more impacted, especially when the face is subject to Yaw and Diag movements, unlike other approaches, which are quite robust to Yaw movements. Pitch and Diag movements challenge all approaches; in their presence, eyebrows and facial edges are generally not detected well.

Regarding facial expressions (Figure 4-B), HG and TCDCN are once again strongly impacted, especially around the eyebrows and the facial contours. DAN and FHR have the same problem but to a lesser extent. On all approaches, expressions of Disgust and Sadness highly interfere with the detection of landmarks around the mouth. This is because lip crease movements are often not well detected.

In the presence of both facial expressions and head pose variations (Figure 4-C), all approaches encounter difficulties at different levels. The most erroneous landmarks are usually located around the eyebrows and the facial contours. Landmarks within the face, especially around the mouth, seem less affected by the combined presence of expressions and head pose variations. The errors reported in the last column (4-C) are similar to the ones obtained in the first two (4-A and 4-B).

4.5. Discussion

The difficulties encountered by facial alignment approaches are mainly related to some expressions (Disgust and Sadness) and to some head pose variations (Pitch and Diag). One reason might be that Disgust and Sadness expressions, Pitch and Diag head movements are less present in the datasets considered for training (see Table 1). Besides, the above expressions and head movements are intrinsically complex. Disgust and Sadness present a wide range of activation patterns and intensities resulting in a very wide set of instances that make converge difficult. Pitch and Diag movements corresponds to one or several out-of-plane rotations where landmark agglutinations interfere with the localization process.

Concerning the face alignment approaches analyzed in this study, the static approaches with temporal constraints represented by FHR and SBR show an ability to reduce the FR, but do not always increase the accuracy significantly. Relying on the use of style-aggregated images to deal with environmental changes, SAN shows its effectiveness in the presence of both head pose variations and facial expressions.

The combination of expressions and head movements, which is close to the conditions of natural interactions, increases the difficulty of landmark detection. The results obtained show that current approaches are not robust enough to detect facial landmarks in complex situations.

However, although some facial landmarks are not well detected, it is important to consider the importance of these facial landmarks in characterizing a face. Indeed, it may not be necessary to accurately detect every facial landmark on a face to properly characterize it, depending on the target application. For instance, landmarks

around the mouth or the eyebrows seem more representative of some facial expressions than landmarks at the contours of the face. To answer this question, it is necessary to analyze the impact of landmark detection errors on subsequent tasks, in our case facial expression recognition.

5. Impact of face alignment on facial expression recognition

In this section, we evaluate how errors in landmark detection impact one subsequent task: expression recognition. First, adopting an expression preserving landmark-based registration approach, we measure the impact of face alignment errors on expression recognition. We consider to experimental settings. In the first one, we train classifiers on frontal faces and then we provide aligned faces for testing. In the second one, we train classifiers on aligned faces and use also aligned faces for testing. Hence, we can assess the impact of landmark detection quality in two common settings encountered when facial expression recognition is conducted in presence of head pose variations.

5.1. Training the classifier with frontal faces

In this experiment, the recordings of SNaP-2DFe produced through the helmet camera (see first row in Fig. 3, page 6) are used to train a frontal facial expression model. Based on the outputs of the face alignment models, we perform face registration on the sequences recorded by the static camera (see second row in Fig. 3, page 6) in order to correct head pose variations and obtain frontal faces. The sequences from the static camera are registered using the 3D face registration approach and are used as the test set. Then, we evaluate the ability of face alignment approaches to provide reliable input for face registration, with expression recognition as the final objective.

We compare the performance of several facial alignment approaches for face registration according to the ability of different descriptors (LMP [4] and LBP [20]) to characterize facial expressions. The use of a motion-based descriptor like LMP makes it possible to highlight the ability of face alignment approaches to provide stable landmarks over two successive images.

The results are given in Table 6. Each accuracy value is computed with SVM, using a ten-fold cross validation protocol on seven expressions (i.e., Anger, Disgust, Fear, Happiness, Neutral, Sadness and Surprise) and on six head pose variations (Static, Tx, Roll, Yaw, Pitch and Diag).

Table 6. Comparison of facial expression performances based on different face alignment approaches.

	Original face		Registered face						
Method	Helmet	Static	Ground Truth	HG	TCDCN	DAN	FHR	SBR	SAN
LBP (texture-based)	75.6%	49.8%	70.3%	67.9%	66.0%	60.0%	67.9%	65.7%	66.3%
LMP (motion-based)	86.0%	47.9%	62.9%	46.8%	42.1%	43.0%	55.2%	57.3%	60.0%

The results in Table 6 show that temporal descriptors are performing well for facial expression analysis in the absence of head movements (helmet camera). However, a drastic fall in performance on the original data from the static camera is observed. In this context, the approaches yield worse results because they suffer from the presence of head pose variations (e.g., Roll, Pitch, Yaw) and large displacements (e.g., Tx, Diag). It is important,

therefore, to ensure that the landmarks driving the registration process are stable over time in order to maintain the benefits of the dynamic descriptors.

The use of registration based on the landmarks of the ground truth improves significantly the performance of facial expression recognition. Still, this remains insufficient compared to the performances obtained on the helmet camera. This is also the case when using landmarks calculated by recent face alignment approaches. Although most approaches achieve performance that tends to be close to the ground truth concerning the texture-based descriptor, there is still a significant difference between the ground truth and the results obtained by face alignment approaches with the motion-based descriptor. This difference can be explained by the fact that the landmarks of these approaches are less stable over short sequences than those provided by the ground truth, which results in temporal artifacts in the reconstruction of the facial movement.

In the light of these results, each face alignment approach tends to increase the performances of facial expression recognition. However, the recognition results still remain lower. In the experiments reported in this section, one bias corresponds to the fact that the training was performed on the original data from the helmet camera, that does not suffer from registration artifacts as the test data does. In the following, we analyze the impact of each alignment approach in an experimental setting where the training data is collected from the static camera and a landmark-based registration is performed prior to training.

5.2. Training the classifier with registered faces

In this evaluation, we consider the registered faces for training and testing the classifier. Hence, we evaluate whether the face alignment and face registration steps preserve distinctive features related to the expressions themselves.

5.2.1. Impact of head movements on recognition accuracy

Table 7 contains the difference (ΔAcc) between the accuracy value obtained using the landmarks of the ground truth and the accuracy value obtained using the landmarks of the different face alignment approaches. The accuracy is computed using a ten-fold cross validation protocol on each 6 movement-based subsets. The left-hand part of Table 7 concerns in-plane head motions (e.g., Static, Tx, and Roll); the right-hand part of Table 7 concerns head motions with out-of-plane rotations (e.g., Yaw, Pitch, and Diag).

The performances of the texture-based descriptor (LBP) following a registration based on the landmarks of the ground truth or those of the face alignment models are relatively similar on in-plane variations (Static, Tx, and Roll), except for HG and TCDCN. In the presence of out-of-plane movements (Yaw, Pitch, and Diag), all face alignment approaches behave rather similarly and are relatively close to the performance obtained with the ground truth landmarks, except for the Pitch movement.

For the motion-based descriptor (LMP), the difference between the ground truth and other face alignment approaches is more significant. This is mainly due to the fact that landmarks are not stable over consecutive images, which produces motion discontinuities. However, FHR, SBR, and SAN achieve performances closer

Table 7. ΔAcc of recognition rate between registered faces based on GT and others on varying head poses.

Face alignment	Descriptor	Static	Tx	Roll	Yaw	Pitch	Diag	Average
HG	LBP	-5.7%	-1.9%	+1.0%	+2.9%	-17.2%	0.0%	-3.5%
	LMP	-7.6%	-22.9%	-19.1%	-3.9%	-12.4%	-9.5%	-12.6%
TCDCN	LBP	-4.7%	-4.7%	-1.9%	-1.0%	-3.8%	+0.9%	-2.5%
	LMP	-15.3%	-12.4%	-18.1%	-15.3%	-14.3%	-16.2%	-15.3%
DAN	LBP	0.0%	+1.0%	-2.8%	-1.9%	-12.4%	-6.7%	-3.8%
	LMP	-16.2%	-13.4%	-15.3%	0.0%	-16.2%	-10.5%	-11.9%
FHR	LBP	0.0%	-2.8%	0.0%	+4.8%	-14.3%	-5.8%	-3.0%
	LMP	-6.7%	-7.7%	-7.7%	+5.9%	-14.3%	+5.7%	-4.1%
SBR	LBP	+1.9%	-0.9%	-3.8%	-1.9%	-15.3%	-2.9%	-3.8%
	LMP	-5.7%	-9.6%	-1.0%	+2.8%	-15.2%	-11.4%	-6.7%
SAN	LBP	-3.8%	-4.7%	-1.9%	-1.9%	-14.3%	-2.9%	-4.9%
	LMP	+2.8%	-0.1%	-3.8%	+11.4%	-10.5%	-2.9%	-0.5%
Average	LBP	-2.1%	-2.3%	-1.6%	0.2%	-12.9%	-2.9%	-3.6%
	LMP	-8.1%	-11.0%	-10.8%	+0.2%	-13.8%	-7.5%	-8.5%

to the ground truth, which may indicate that these approaches tend to be more stable over time; FHR and SBR were designed to be stable over time, FHR by including a temporal smoothing term in its loss, and SBR by being trained to mimic the KLT tracker. Whether there is movement in or out of the plane, HG, TCDCN, and DAN display weaknesses, with the exception of the Yaw movement for DAN. FHR, SBR and SAN perform relatively similarly to the ground truth in the presence of in-plane movements. However, when there are out-of-plane movements, the difference is larger. It is interesting to see that FHR and SAN tend to perform better than the ground truth on Yaw and Diag. This may be explained by the fact that the approach used to register the face is trained on landmarks that are more similar to those provided by these approaches than those provided by the ground truth.

5.2.2. Impact of landmark quality per expression

In this experiment, we analyze the impact of landmark quality on the performance for each facial expression. Table 8 provides the difference between the accuracy (Δ Acc.) obtained using the landmarks of the ground truth and the ones obtained using the outputs of the different face alignment approaches. As for the previous evaluation, only the images from the static camera are used for training and testing. For each expression, the accuracy is calculated by considering one expression against all the others. Ten-fold cross validation is used for the evaluation on each of the six expression subsets.

Table 8 shows an average difference of -1.0% in terms of accuracy compared to the ground truth for the texture-based descriptor (LBP) and -1.40% for the motion-based descriptor (LMP). The difference is larger for LMP because the stabilization of facial landmarks during the sequence induces movement discontinuities that tend to reduce its performance. This is reflected in the results obtained by SBR and FHR, both based on solutions that improve the stability of the landmarks by adding temporal constraints.

Expressions involving significant geometric deformations (Happiness, Surprise, and Disgust) tend to worsen performances. Under these conditions, a landmark detection error has a larger impact on facial registration and tends to deform the initial facial geometry. For Anger and Sadness, TCDCN and HG provide the worst

Table 8. ΔAcc between registered faces based on GT and others on varying facial expressions.

Face alignment	Descriptor	Happiness	Fear	Surprise	Anger	Disgust	Sadness	Average
HG	LBP	-0.2%	+0.8%	-1.3%	-0.7%	-1.5%	+1.1%	-0.3%
	LMP	-4.3%	+0.2%	-1.1%	-3.5%	-4.4%	-0.7%	-2.3%
TCDCN	LBP	+0.1%	+2.4%	-0.7%	-0.9%	-1.9%	-2.8%	-0.6%
	LMP	-4.0%	0.0%	-1.4%	-2.9%	-4.3%	-1.7%	-2.4%
DAN	LBP	-1.9%	+0.6%	-3.3%	-0.9%	-3.6%	-1.0%	-1.7%
	LMP	-5.5%	+0.2%	-2.2%	-1.8%	-5.2%	-1.8%	-2.7%
FHR	LBP	-1.2%	+0.4%	-2.2%	-1.5%	-2.1%	-1.2%	-1.3%
	LMP	-1.1%	0.0%	+1.5%	+0.9%	-1.3%	+0.2%	+0.03%
SBR	LBP	-1.5%	+0.4%	-3.1%	-0.4%	-1.5%	+0.7%	-0.9%
	LMP	-3.3%	+0.2%	+0.6%	-0.9%	-0.9%	+0.4%	-0.7%
SAN	LBP	-1.0%	+0.6%	-3.1%	-0.6%	-1.3%	-1.9%	-1.2%
	LMP	-2.5%	0.0%	0.0%	-0.3%	-0.7%	+1.1%	-0.4%
Average	LBP	-0.9%	+0.9 %	-2.3%	-0.8 %	-1.9%	-0.9	-1.0%
	LMP	-3.5 %	+0.1%	-0.4 %	-1.4 %	-2.8 %	-0.4	-1.4%

performance. This is mainly due to their difficulties in tracking lip deformations in the presence of these expressions. It has an strong impact on the resulting registration.

Although some face alignment approaches detect landmarks with a less-than-perfect accuracy, they tend to yield better performance than the ground truth for some expressions. Considering the synthesis presented in Table 9, HG and TCDCN perform very well when using static descriptors such as LBP. However, results decreases for this two face alignment solutions when dynamic descriptors are used. FHT, SAN, and SBR perform better when face alignment is required to be stable and robust in time.

Table 9. Synthesis of face alignment performances (AUC) and facial expression recognition (FER) accuracy when computed using LBP (static) or LMP (dynamic) features. The rank for each face alignment method is provided with regard to the different tasks.

	Alignment		FER LBP		FER LMP	
	AUC	Rank	Acc.	Rank	Acc.	Rank
HG	52.94%	6	-0.30%	1	-2.30%	4
TCDCN	52.97%	5	-0.63%	2	-2.38%	5
DAN	71.77%	4	-1.68%	6	-2.72%	6
FHR	71.84%	3	-1.30%	5	0.03%	1
SBR	73.76%	1	-0.90%	3	-0.65%	3
SAN	73.38%	2	-1.22%	4	-0.40%	2

5.3. Impact per landmark

In this evaluation, we assess the impact of errors in facial landmarks detection on facial expression recognition using the linear regression analysis. For each of the 68 landmarks, we calculate a regression coefficient based on the detection error of facial landmarks on all face alignment approaches relative to the accuracy of facial expression recognition. The importance of a facial landmark is assessed according to the rank of its regression coefficient in relation to other facial landmarks. Figure 5 shows the importance of facial landmarks to preserve facial geometry according to head pose variations (first line) and facial expressions (second line).

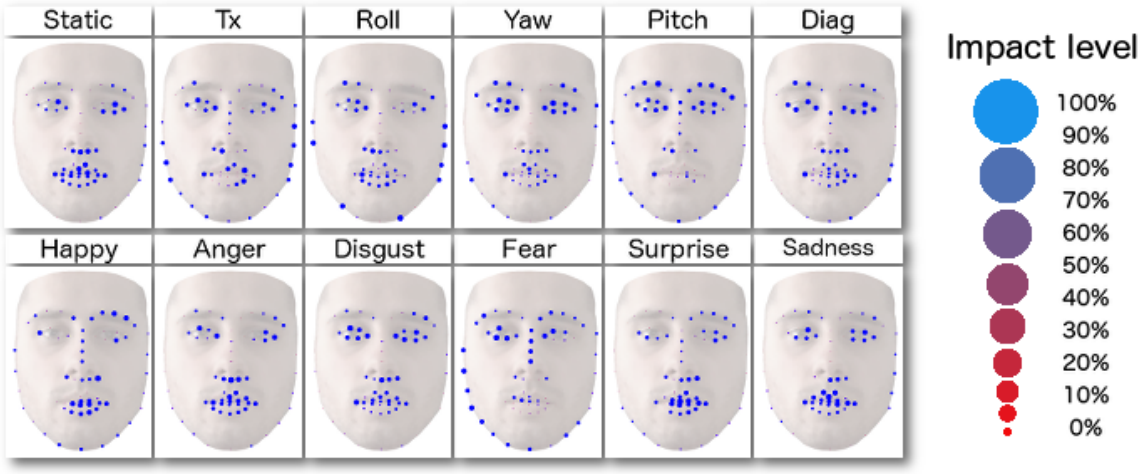

Figure 5. Importance of landmarks in the preservation of facial geometry according to head poses and facial expressions.

Concerning the impact of facial landmarks in the presence of head pose variations, it is interesting to note that facial landmarks on the contours of the face and at the nasal ridge are not too significant for the majority of poses. This is probably due to the fact that a triangulation between the position of the eyes and the center of the nose is sufficient to correctly estimate the facial pose and register the face. When the face is static, the important facial landmarks are mainly located on the lips. In the presence of movement, the important facial landmarks tend to be located on different regions inside the face. Concerning out-of-plane movement, it is interesting to see that facial landmarks at the eyebrow and eye level are very important to avoid deforming the face for the Pitch movement. It can be noted that this is different for the Yaw and Diag movements, where the visible face part is as important as the occluded face part.

Concerning facial expressions, we also observe that the facial landmarks located on the contours of the face and at the nasal ridge are mostly not significant. On all expressions, facial landmarks at the level of the mouth are very important except for the Fear expression, in which eyebrows are more dominant.

Overall, the most important facial landmarks to register the head pose while maintaining facial expressions are located at the eye and mouth levels. Although facial landmarks at the nasal ridge are not too important, it is interesting to note that facial landmarks at the lower nose are still strongly present in all heatmaps. These facial landmarks are probably what allows the facial registration model to abstract itself from the other facial landmarks characterizing the rigid regions of the face (contours, nasal ridge).

5.4. Discussion

The analysis of facial expressions in the presence of head pose variations generally requires the use of a face registration technique to bring the face into an ideal setting. Face registration techniques use facial landmarks provided by face alignment approaches to estimate the correct transformation to be applied to the face to correct the head pose. This requires that the facial landmarks are correctly detected in order to avoid registration errors.

In this evaluation, we studied the importance of facial landmarks in the preservation of facial geometry according to head pose variations and facial expressions. Based on the various evaluations, we have shown that recent face alignment approaches have various impact levels on the preservation of facial geometry during face registration. The impact varies according to the complexity of the face to be characterized. Indeed, the head pose variations challenge more the face alignment approaches, which also has an impact on the recognition accuracy obtained when characterizing the facial expressions on the registered faces.

Regarding face alignment approaches, six approaches are considered: DAN, FHR, HG, SAN, SBR, and TCDCN. SBR and FHR have the advantage of leveraging temporal information. This has a positive impact on motion-based facial expression analysis because more stable predictions are obtained between two successive images. SAN has the particularity of better characterizing facial landmarks inside the face through the use of style-aggregated images that are more robust to environmental changes.

The impact of different landmarks on the preservation of facial geometry shows that some facial regions are more important. It is interesting to draw a parallel between landmark detection error rates and landmark impact for the preservation of facial geometry. In view of the results obtained, it is more interesting to favour face alignment approaches that correctly detect landmarks on the mouth and eyes rather than those focusing on the outer contours of the face. Indeed, the position of the eyes, lower nose and mouth seems sufficient to correct the head pose variations. Mouth landmarks are also very important to characterize the different facial expressions.

6. Conclusion

In this study we first addressed the questions of the quality of facial landmark detection in the simultaneous presence of head pose variations and facial expressions. Then, we have studied in the importance of properly detecting facial landmarks for underlying tasks such as facial expression analysis.

The results obtained show that face alignment approaches tend to become increasingly robust in the presence of head pose variations and facial expressions. However, some conditions still challenge these approaches. Among these conditions, we can distinguish the out-of-the-plane rotations (especially Pitch and a combination of Pitch and Yaw) and expressions that involve some complex mouth movements such as Disgust and Sadness.

By studying the importance of the different facial landmarks to correct head pose, we have shown that facial landmarks on the outer contours of the face and on the ridge of the nose are of little importance and that it is more important to correctly detect facial landmarks on the eyebrows, eyes, and mouth.

Based on this assessment, it would be interesting if competitions on face alignment could took into consideration the impact of landmarks on other subsequent tasks such as facial expression analysis. It would make it possible to identify important steps in the training process in order to strengthen the detection of important facial landmarks to the detriment of other, less significant, facial landmarks. In particular, in these evaluations, we have shown that taking into account some temporal information in the training process or using solutions to correct intra-face variations before detection improves the robustness and stabilization of facial landmark detection, and thus improves the quality of the resulting facial registration.

We believe that the development of datasets like SNaP-2DFe can drive the community to design facial alignment approaches that meet the expectations of the users of these approaches and improve the performance of their systems.

1. Deng, J.; Roussos, A.; Chrysos, G.; Ververas, E.; Kotsia, I.; Shen, J.; Zafeiriou, S. The Menpo benchmark for multi-pose 2D and 3D facial landmark localisation and tracking. International Journal of Computer Vision **2018**, pp. 1–26.
2. Chrysos, G.G.; Antonakos, E.; Snape, P.; Asthana, A.; Zafeiriou, S. A comprehensive performance evaluation of deformable face tracking "in-the-wild". International Journal of Computer Vision **2018**, 126, 198–232.
3. Hassner, T.; Harel, S.; Paz, E.; Enbar, R. Effective face frontalization in unconstrained images. Conference on Computer Vision and Pattern Recognition, 2015, pp. 4295–4304.
4. Allaert, B.; Bilasco, I.M.; Djeraba, C. Advanced local motion patterns for macro and micro facial expression recognition. arXiv preprint arXiv:1805.01951 **2018**.
5. Zafeiriou, S.; Trigeorgis, G.; Chrysos, G.; Deng, J.; Shen, J. The Menpo facial landmark localisation challenge: A step towards the solution. Conference on Computer Vision and Pattern Recognition Workshops, 2017.
6. Bulat, A.; Tzimiropoulos, G. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). International Conference on Computer Vision, 2017.
7. Kowalski, M.; Naruniec, J.; Trzcinski, T. Deep alignment network: A convolutional neural network for robust face alignment. Conference on Computer Vision and Pattern Recognition Workshops, 2017.
8. Zhang, Z.; Luo, P.; Loy, C.C.; Tang, X. Learning deep representation for face alignment with auxiliary attributes. IEEE Transactions on Pattern Analysis and Machine Intelligence **2016**, 38, 918–930.
9. Dong, X.; Yan, Y.; Ouyang, W.; Yang, Y. Style aggregated network for facial landmark detection. Conference on Computer Vision and Pattern Recognition, 2018, pp. 379–388.
10. Wang, W.; Tulyakov, S.; Sebe, N. Recurrent convolutional shape regression. IEEE Transactions on Pattern Analysis and Machine Intelligence **2018**, 40, 2569–2582.
11. Liu, H.; Lu, J.; Feng, J.; Zhou, J. Two-stream transformer networks for video-based face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **2018**, 40, 2546–2554.
12. Gu, J.; Yang, X.; De Mello, S.; Kautz, J. Dynamic facial analysis: From Bayesian filtering to recurrent neural network. Conference on Computer Vision and Pattern Recognition, 2017, pp. 1548–1557.
13. Peng, X.; Feris, R.S.; Wang, X.; Metaxas, D.N. RED-Net: A recurrent encoder–decoder network for video-based face alignment. International Journal of Computer Vision **2018**, 126, 1103–1119.
14. Dong, X.; Yu, S.I.; Weng, X.; Wei, S.E.; Yang, Y.; Sheikh, Y. Supervision-by-Registration: An unsupervised approach to improve the precision of facial landmark detectors. Conference on Computer Vision and Pattern Recognition, 2018, pp. 360–368.
15. Tai, Y.; Liang, Y.; Liu, X.; Duan, L.; Li, J.; Wang, C.; Huang, F.; Chen, Y. Towards highly accurate and stable face alignment for high-resolution videos. AAAI Conference on Artificial Intelligence, 2019.
16. Belmonte, R.; Ihaddadene, N.; Tirilly, P.; Bilasco, I.M.; Djeraba, C. Video-based Face Alignment with Local Motion Modeling. IEEE Winter Conference on Applications of Computer Vision **2019**.
17. Tulyakov, S.; Jeni, L.A.; Cohn, J.F.; Sebe, N. Viewpoint-consistent 3D face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence **2018**, 40, 2250–2264.
18. Zhu, X.; Liu, X.; Lei, Z.; Li, S.Z. Face alignment in full pose range: A 3D total solution. IEEE Transactions on Pattern Analysis and Machine Intelligence **2019**, 41, 78–92.
19. Jourabloo, A.; Liu, X. Pose-invariant face alignment via CNN-based dense 3D model fitting. International Journal of Computer Vision **2017**, 124, 187–203.
20. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **2002**, 24, 971–987.
21. Ding, H.; Zhou, S.K.; Chellappa, R. Facenet2expnet: Regularizing a deep face recognition net for expression recognition. International Conference Automatic Face & Gesture Recognition, 2017, pp. 118–126.
22. Bassili, J.N. Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. Journal of Personality and Social Psychology **1979**, 37, 2049–2058.

23. Zhao, G.; Pietikainen, M. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence **2007**, 29.
24. Schaefer, S.; McPhail, T.; Warren, J. Image deformation using moving least squares. ACM Transactions on Graphics (TOG). ACM, 2006, Vol. 25 (3), pp. 533–540.
25. Allaert, B.; Mennesson, J.; Bilasco, I.M.; Djeraba, C. Impact of the face registration techniques on facial expressions recognition. Signal Processing: Image Communication **2018**, 61, 44–53.
26. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. Conference on Computer Vision and Pattern Recognition, 2014, pp. 1867–1874.
27. Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 faces in-the-wild challenge: Database and results. Image and Vision Computing **2016**, 47, 3–18.
28. Le, V.; Brandt, J.; Lin, Z.; Bourdev, L.; Huang, T.S. Interactive facial feature localization. European Conference on Computer Vision, 2012, pp. 679–692.
29. Koestinger, M.; Wohlhart, P.; Roth, P.M.; Bischof, H. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. International Conference on Computer Vision Workshops, 2011, pp. 2144–2151.
30. Burgos-Artizzu, X.P.; Perona, P.; Dollár, P. Robust face landmark estimation under occlusion. International Conference on Computer Vision, 2013, pp. 1513–1520.
31. Zhu, X.; Lei, Z.; Liu, X.; Shi, H.; Li, S.Z. Face alignment across large poses: A 3D solution. Conference on Computer Vision and Pattern Recognition, 2016, pp. 146–155.
32. Zafeiriou, S.; Chrysos, G.G.; Roussos, A.; Ververas, E.; Deng, J.; Trigeorgis, G. The 3D Menpo facial landmark tracking challenge. International Conference on Computer Vision, 2017, pp. 2503–2511.
33. Shen, J.; Zafeiriou, S.; Chrysos, G.G.; Kossaifi, J.; Tzimiropoulos, G.; Pantic, M. The first facial landmark tracking in-the-wild challenge: Benchmark and results. International Conference on Computer Vision Workshops, 2015, pp. 50–58.