Nose pose estimation in the wild and its applications on nose tracking and 3D face alignment

Flávio H. de B. Zavan, Luciano Silva and Olga R. P. Bellon
Imago Research Group
Universidade Federal do Paraná
Curitiba, Paraná, Brazil, 81531-980
Email: {flavio,luciano,olga}@ufpr.br

***Abstract*—An automatic, landmark free SVM-based method for head pose estimation, solely using the nose region, in constrained and unconstrained scenarios, is presented. Using the nose region has advantages over the whole face; it is less likely to be occluded or deformed by facial expressions, and is proven to be highly discriminant in all poses from profile to frontal. The approach, SVM-NosePose, receives a nose region as and classifies it into a discrete set of poses. Estimation favorably compares against state-of-the-art works on six publicly available datasets. Three applications are derived from the proposed methodology: 1) the original inclusion of a head pose score for face quality estimation for initializing a nose tracker, leading to higher accuracy; 2) 3D face alignment in the wild using only the nose pose, enabling consistent estimates even in challenging scenarios; and 3) multipose action unit detection and intensity estimation for facial images in the wild.**

I. Introduction

Head pose estimation is defined as determining at least one of the three parameters that configures the face relative to its three degrees of freedom, yaw, pitch and roll, and the camera [1]. It has direct applications in other computer vision problems, including gaze estimation [2], face quality assessment [3], face recognition [4], facial landmark detection [5], automatic affect analysis in infants [6] and face frontalization [7].

Although the whole face is traditionally used for estimating the head pose [1] and the scientific community has been showing great interest in depth images [8], [9], this work focuses on unconstrained environments, where extreme head poses are common and no specific sensor is used. Under such conditions, face detection and pose estimation are considered difficult problems [10].

SVM has been used to classify the gradient information of the nose region into a discrete set of angles [11], but it was only applied on one controlled dataset. For estimating the pose in the wild using the whole face, Peng *et al.* [12] perform manifold analysis, while Demirkus *et al.* [13] aggregate probability density functions, estimated from facial features, based on temporal information.

This work proves that the nose can be successfully used for estimating the head pose in both constrained and unconstrained environments, in a variety of datasets. The use of the nose for face processing has previously been shown efficient [11], [14], [15], as it has multiple desirable qualities. Unlike the eyes and ears, it is visible even in profile faces; unlike the mouth, it cannot be easily deformed by speech and expressions; it is also less likely to be partly occluded by accessories and facial traits, such as sunglasses and beards, when compared to using the whole face. In addition, the head pose is needed for many computer vision applications.

SVM-NosePose was developed and classifies the head pose into a discrete set of angles using Support Vector Machines (SVM), trained with the output of the Local Gradient Increasing Pattern (LGIP) filter [16]. SVM-NosePose is coupled with the state-of-the-art Faster R-CNN detector [17] for making the method fully automatic.

An evolution of the SVM-NosePose method, based on CNNs (Convolutional Neural Networks), and three direct applications are also discussed: 1) Enhancing an existing face quality estimator [18] by providing a head pose score and using it for initializing a nose tracker; 2) Generating consistent estimations for landmark free 3D face alignment in the wild under extreme poses and facial expressions; 3) Additionally, NosePose is used as a regularization term in a CNN for detecting action units and estimating their intensity.

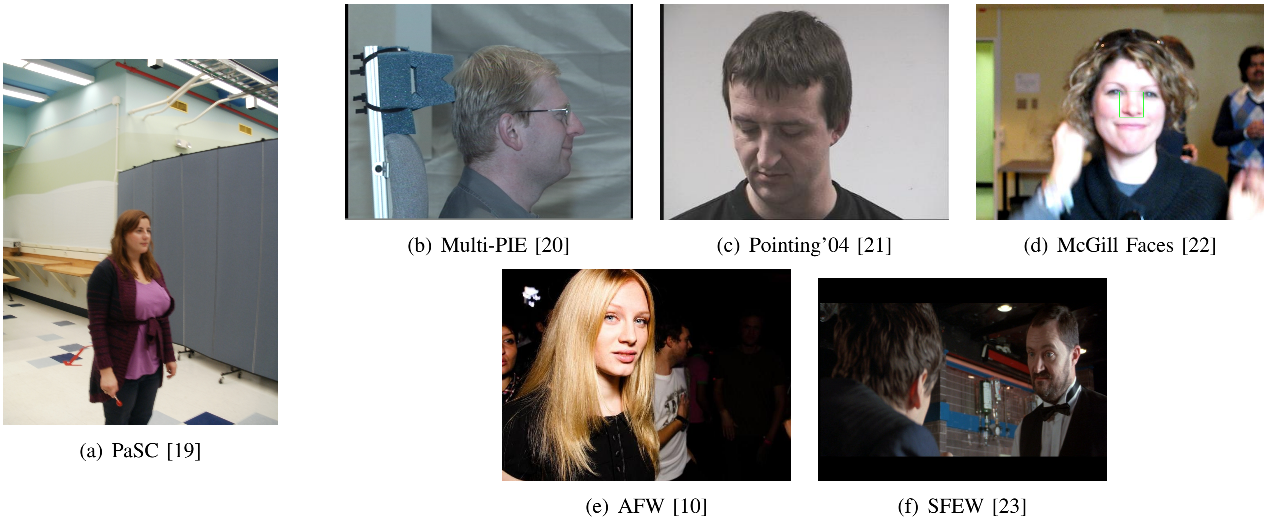
SVM-NosePose's performance is evaluated on six publicly available datasets. Two of which were built using images acquired in controlled environments and four which include images from unconstrained environments. Examples from both kinds are shown in Figure 1. When possible and appropriate, SVM-NosePose is compared against the state-of-the-art, enabling and in-depth performance evaluation to be carried.

This paper is organized as follows: Section 2 presents the head pose estimation method and its individual steps; Section 3 discusses the obtained results on all datasets for both the whole method and its individual steps; Section 4 introduces SVM-NosePose's derived works and direct contributions; finally, Section 5 concludes with the final remarks.

II. SVM-NosePose

SVM-NosePose can be divided into three main steps: detection, feature extraction and classification. The state-of-the-art generic object detector, Faster R-CNN [24], trained specifically for the nose detection task, composes the first step. This method is an evolution of Fast R-CNN [25], it addresses its proposal generation performance by introducing

This paper is based on the work of a M.Sc. dissertation



(a) PaSC [19]

(b) Multi-PIE [20]

(c) Pointing'04 [21]

(d) McGill Faces [22]

(e) AFW [10]

(f) SFEW [23]

Fig. 1: Example images acquired in both constrained (b and c) and unconstrained (a, d, e and f) environments

the concept of Region Proposal Networks, which are able to create detection proposals in near real-time. Each possibility is subsequently evaluated by existing deep network models before outputting the final detections. SVM-NosePose normalizes all nose regions to 54x60 pixels, an empirically determined size, before the next step.

When defining the feature extraction step, using Pawelczyk and Kawulok's [11] as basis, multiple descriptors were tested. Results show that binary patterns accurately describe the nose region for estimating the head pose, including LBP [26], LGP [27] and LGIP [16]. Further experimentation indicate that the best performance can be achieved with subregion LGIP histograms. By subdividing the nose region, some spatial information is allowed to be present while also allowing some variation to occur. The exact number of subregions is determined during training and varies with each dataset. An exhaustive search is performed with all perfect squares from 1 to 100, the one that maximizes the total accuracy is chosen. The descriptor selection, experimentation and the adjustment of the number of regions is detailed in [28].

Classification is performed using SVM. The LGIP filter [16] is applied on the detected nose region, which is subsequently divided into subregions. The histogram of each is calculated and concatenated, forming the feature vector. A SVM with a RFB kernel is trained with 10-fold cross validation for classifying the features into a discrete set of poses. The granularity of the existing classes depends on each dataset, varying from 15 to 45 degrees. This process is presented as a diagram in Figure 2.

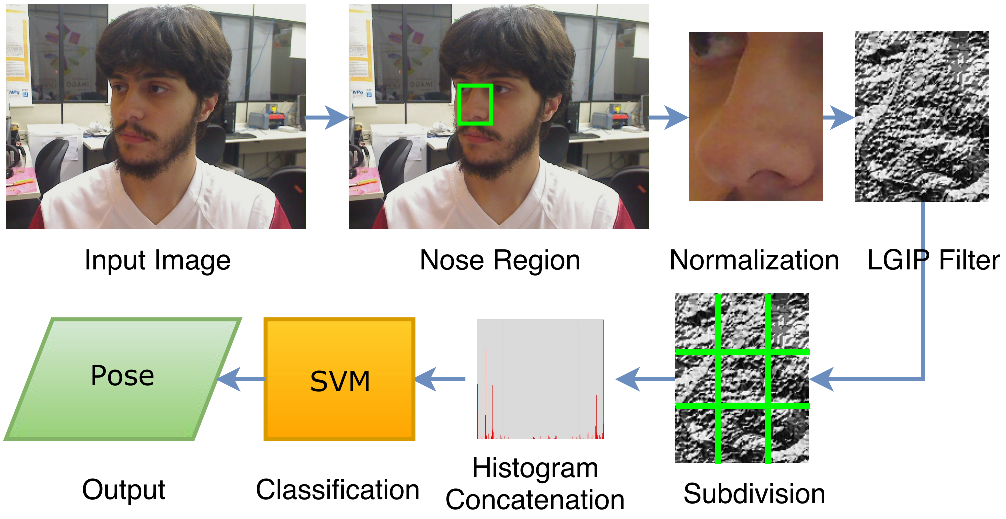


Fig. 2: SVM-NosePose diagram [29]

III. Experimental Results

SVM-NosePose was evaluated on six publicly available datasets. Two of which were acquired in controlled environments, Multi-PIE [20] and Pointing'04 [21], and four in unconstrained scenarios, McGillFaces [22], SFEW [23], PaSC [19] and AFW [10].

Each dataset encompasses different difficulties and challenges (Figure 1), allowing for a thorough evaluation. It is important to note that extreme poses are underrepresented in all unconstrained datasets in this work. Due to their nature, controlled datasets tend to have an equal distribution of the possible variations in the head yaw. However, many common applications for in-the-wild face processing tend to favor near frontal poses. These differences are well represented in the different datasets (Figure 3).

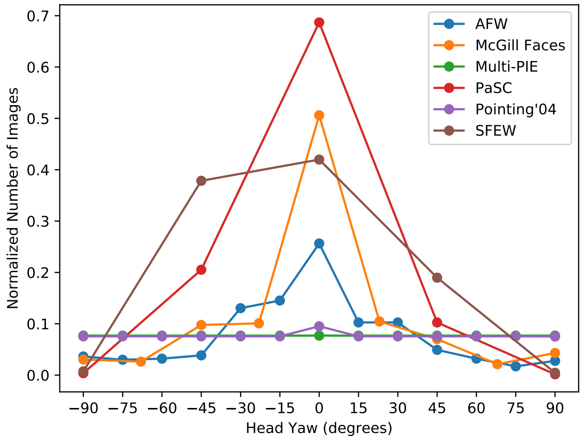


Fig. 3: Head yaw distribution on all datasets

Because SVM-NosePose depends on Faster R-CNN [24] to be fully automatic, the detector is evaluated separately on all datasets in order to isolate, determine and understand the limitations of the methodology.

A. Nose Detection

Nose detection performance is first evaluated in isolation. For each dataset, Faster R-CNN [24] is trained using all default parameters for detecting the nose region. In order to assess the performance, Hoover *et al.*'s intersection coefficient [30] (Equation 1) between the detection region and the ground-truth annotation is used as metric.

$$\begin{split}ic(A,B) = min(area(intersection(A,B))/area(A),\\ area(intersection(A,B))/area(B))\end{split} \tag{1}$$

Faster R-CNN is able to detect the nose region with great performance on Multi-PIE, Pointing'04, McGillFaces and SFEW. However, performance is degraded on PaSC and AFW. AFW contains a limited number of images, which directly influences Faster R-CNN's ability to learn. PaSC contains a significant number of low quality and low resolution faces and noses, including completely blurred regions. Unlike the other datasets, the subjects in PaSC represent only a small area of the images, which is incompatible with the default training parameters. To evaluate the effects of the size of the image, a second nose detector was trained on PaSC, using only the cropped face regions as input, generating considerably lower false positive rates.

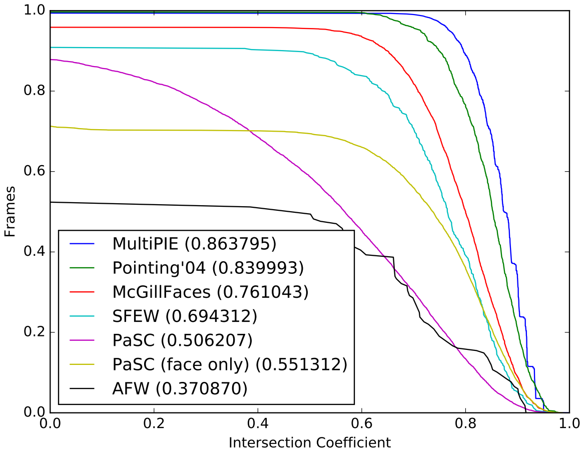


Fig. 4: Intersection Coefficient curves on all datasets. The number in parenthesis is the area under the curve [28]

The achieved accuracy and false positive rates are available in Table I. Similarly, Figure 4 presents the results in curves after using Faster R-CNN's internal detection score for filtering down to one detection for each subject in the image.

TABLE I: Percentage of images where the intersection coefficient of the detected nose region and the annotated ground-truth is at least 0.5 and the amount of false positives [28]

Dataset	Accuracy	False Positives
Multi-PIE	99.66%	15.52%
Pointing'04	99.89%	29.38%
McGillFaces	97.21%	22.56%
SFEW	90.86%	14.21%
PaSC	82.89%	92.51%
PaSC (face only)	73.66%	47.97%
AFW	51.19%	79.18%

B. Pose Estimation

SVM-NosePose was evaluated using both the detected nose regions and ground-truth (GT) annotations, for individually assessing the performance of the pose estimation step. Two evaluation protocols are used when applicable: *strict*, only correct classifications are considered hits; and *weak*, when off-by-one classifications are considered hits.

Multi-PIE [20] contains images of over 300 subjects in a controlled scenario, including variations in illumination and facial expression. Each scene is captured simultaneously with 15 cameras, two overhead sensors and 13 representing different head poses from -90 to 90 degrees along the yaw axis, in steps of 15. SVM-Nosepose was trained with 10,000 images and tested on 355,900. All nose regions were annotated semi automatically due to the controlled nature of the dataset. Despite the large number of classes when estimating the pose on the yaw axis, the controlled environment favors high hit rates with both protocols when SVM-NosePose is tested with the ground-truth nose regions (Table II). When the nose is detected automatically, performance degrades only with the strict protocol.

Pointing'04 [21] includes 2,790 images of 15 subjects acquired in a controlled environment. The head pose varies on both yaw and pitch axes from -90 to 90 degrees, totalling 13 classes on the yaw axis and nine on the pitch axis. A total of 1,842 images were used for training and 836, for testing. Results are presented for each axis in Table II. When compared to those achieved on Multi-PIE [20], the strict hit rate is noticeably lower, as the annotation is not perfect [11].

McGillFaces [22] is composed of videos of 60 subjects with a total of 18,000 frames. Challenges include variations in illumination, facial expressions, variations in scale, translation and head pose. However, only of portion of these images are publicly available and only a smaller portion include head yaw annotations (9 classes). The nose region was manually annotated and 3208 images were used for training and 3457, for testing. After inspecting the annotations, inconsistencies were found and a filtered version of the dataset was a generated [29], containing only the images with reliable annotation. The filtered version includes 2475 images for training and 2854 for testing. Results for both are presented in Table III. The effects of the filter are clear in the results, with a significant increase in accuracy.

The Static Facial Expressions in the Wild dataset (SFEW) [23] is collection of movie frames, encompassing multiple different unconstrained scenarios. It contains a total of 1,700 images, subdivided into three subsets. Both nose regions and head poses (5 classes) were manually annotated. When evaluating SVM-NosePose, the training and validation subsets were merged for training (957 images) and the testing subset was used for testing (372 images). Positive results were obtained despite the adversities in the input images (Table III).

The Point and Shoot Challenge dataset (PaSC) [19] contains multiple still images and videos of numerous subjects performing common actions in different indoor and outdoor environments. All still images containing a nose were annotated with both the nose region and head yaw (5 classes). A new subset division of the dataset was used, as the original was conceived for face recognition, containing overlapping subjects, and is not suited for the head pose estimation problem. This process resulted in a total of 5784 training and 6243 testing images. Results on this dataset indicate SVM-NosePose's robustness in unconstrained scenarios (Table III).

While the Annotated Faces in the Wild (AFW) dataset [10] is smaller than all other datasets, with only 205 images and 468 annotated subjects, it is the most challenging. Annotations are provided with a precision of 15 degrees, totalling 13 classes from -90 to 90 in the yaw axis. The nose regions were manually annotated. To train SVM-NosePose on this dataset, 168 subjects are augmented 14-fold, by flipping and rotating, such that 2352 regions are fed to SVM and the remaining 300 are tested on. Performance is similar to the state-of-the-art when the ground-truth nose regions are present. However, due to Faster R-CNN's nature [24], which requires more images for performing efficient detection, performance degrades significantly using the detected regions (Table III).

TABLE II: Head pose estimation performance on controlled datasets

Method	Evaluation	Multi-PIE	Pointing'04 Yaw	Pointing'04 Pitch
SVM-NosePose GT	Strict	94.13%	61.36%	58.49%
SVM-NosePose GT	Weak	99.31%	95.69%	94.50%
SVM-NosePose	Strict	76.67%	45.53%	57.42%
SVM-NosePose	Weak	97.13%	83.96%	92.82%
[10]	Strict	91.40%	–	–
[10]	Weak	99.99%	–	–
[13]	Strict	94.46%	–	–
[11]	Strict	–	56.99%	47.91%
[11]	Weak	–	93.41%	77.80%

On the controlled datasets, SVM-NosePose achieves accuracy rates higher or similar to those using the whole face found in the literature. A decrease in performance can be noticed on the challenging, in the wild, datasets, however, SVM-NosePose still achieves compatible accuracy rates, confirming its competitiveness in such scenarios.

When coupled with the automatic nose detector (Faster R-CNN), performance degrades due to the variations in the detected nose regions, particularly in the smaller datasets where less training data is available, such as AFW. However, the method is still suitable for direct non-trivial applications, discussed in the next section.

TABLE III: Head pose estimation performance on unconstrained datasets

Method	Evaluation	McGillFaces	Filtered McGillFaces	SFEW	PaSC	AFW
SVM-NosePose GT	Strict	59.24%	70.71%	83.60%	86.91%	44.71%
SVM-NosePose GT	Weak	83.34%	92.68%	–	–	81.26%
SVM-NosePose	Strict	–	55.08%	66.67%	75.65%	7.14%
SVM-NosePose	Weak	–	82.98%	–	–	31.55%
[10]	Weak	–	–	–	–	81.00%
[13] (18,000 images)	Strict	79.02%	–	–	–	–

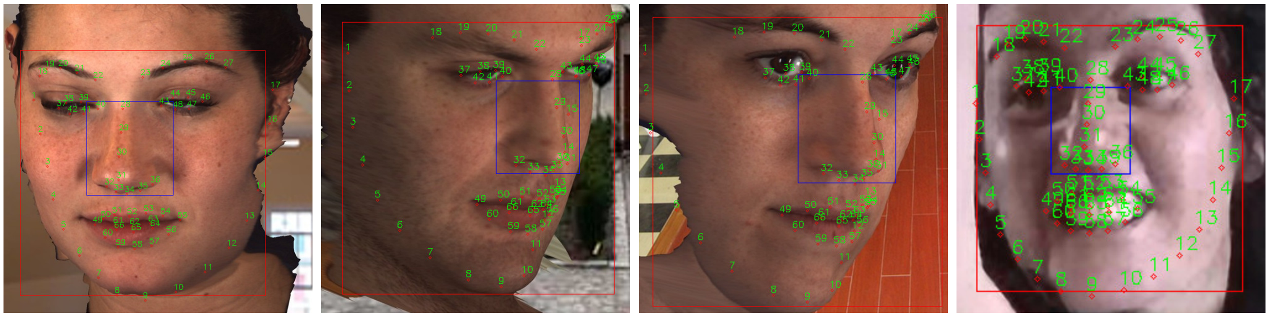


Fig. 5: Example landmark estimations on the 3DFAW challenge [31]

IV. Contributions

Multiple contributions were derived from SVM-NosePose. The ability to accurately estimate the head pose in unconstrained environments is of direct use for other face processing challenges.

An original head pose quality measure was conceived and integrated as an extra feature into an existing face quality estimator [18]. This measure is derived directly from SVM-NosePose's discrete pose output. The value generated by the measurements depends on the problem it is being applied to. In some situations, it can be more beneficial to have frontal faces (*e.g.* face recognition), while other times, different poses may contain more useful information (*e.g.* 3D face reconstruction). This quality measurement is applied for selecting a frame and initializing a state-of-the-art nose tracker. Tests on the 300VW dataset [32] indicate an increase in accuracy when compared to the baseline initialization with the first frame [33]. Temporal information is useful for locating the nose region in difficult environments, where detection methods fail to adequately locate the nose in all frames.

A Convolutional Neural Network (CNN) pose estimator, CNN-NosePose, was derived from SVM-NosePose [29]. This new approach takes the same nose regions as input, but replaces the feature extraction and classification steps, such that they are learned by the CNN, which uses an architecture optimized for estimating the head pose. The trained network eliminates the need for manually providing the number of subregions and is able to better handle the regions detected by Faster R-CNN when compared to SVM-NosePose. An advantage provided by CNNs is the ability to fully utilize and learn from large datasets, favoring success in difficult scenarios with abundant data. Its overall accuracy under such conditions outperforms SVM-NosePose and the state-of-the-art, particularly when a large number of images is available.

As part of the 3D Face Alignment in the Wild challenge (3DFAW) [34], CNN-NosePose was used for estimating the pose in difficult scenarios, *e.g.* facial expressions and extreme head poses. The calculated pose, nose size and location are used for translating, rotating and scaling a generic 3D landmark model, resulting in coherent estimations of the position of the landmarks even in difficult cases [31]. This methodology was also tested with SVM-NosePose and similar results were achieved, due to the straightforwardness in locating the nose in the challenge's images [28]. Two example alignment results can be seen in Figure 5. A further optimization of the position of these landmark estimations was presented by Silva [35], who combined them with a state-of-the-art landmark estimator to produce accurate estimations in both 2D and 3D.

Batista *et al.* [36] proposed AUMPNet, a single CNN architecture for simultaneous facial Action Unit (AU) detection and intensity estimation under multiple head poses. CNN-NosePose's fully connected layers were used aside of the main architecture fully connected layers for handling head pose variations. Thus, the whole architecture was optimized using a multitask loss composed of AU detection, AU intensity regression and head pose estimation. This optimizes AUMPNet's ability to learn better representations under multiple head pose variations.

V. Final Remarks

A fully automatic landmark free method for estimating the head pose in challenging scenarios using only the nose region was presented. It was shown, through experiments where SVM-NosePose outperforms the state-of-the-art, that the nose is a suitable candidate for face processing in the wild. SVM-NosePose's degraded performance when fed with detected nose regions was addressed in derived work [29]. As a confirmation of the method's efficiency and efficacy, an evolution of the work presented [29] was produced and three direct applications of the NosePose methodology [31], [33], [35], [36] were conceived and successfully tested. In future work, CNN-NosePose can be further optimized and applied for solving different face processing problems, including gender estimation, 3D face reconstruction and face recognition. Some of these possibilities are explored and are included in an extended paper that has been submitted to a high impact journal.

Acknowledgment

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

References

[1] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607–626, 2009.

[2] E. D. Guestrin and M. Eizenman, "General theory of remote gaze estimation using the pupil center and corneal reflections," *Biomedical Engineering, IEEE Transactions on*, vol. 53, no. 6, pp. 1124–1133, 2006.

[3] Y. Lee, P. Phillips, J. Filliben, J. Beveridge, and H. Zhang, "Generalizing face quality and factor measures to video," in *Biometrics (IJCB), IEEE International Joint Conference on*, 2014, pp. 1–8.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[5] Z. Zhang, P. Luo, C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision (ECCV)*, ser. Lecture Notes in Computer Science, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Springer International Publishing, 2014, vol. 8694, pp. 94–108.

[6] Z. Hammal, J. F. Cohn, C. Heike, and M. L. Speltz, "Automatic measurement of head and facial movement for analysis and detection of infants' positive and negative affect," *Frontiers in ICT*, vol. 2, p. 21, 2015.

[7] T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[8] S. Tulyakov, R.-L. Vieriu, S. Semeniuta, and N. Sebe, "Robust real-time extreme head pose estimation," in *Pattern Recognition (ICPR), International Conference on*, 2014, pp. 2263–2268.

[9] C. Papazov, T. Marks, and M. Jones, "Real-time 3d head pose and facial landmark estimation from depth images using triangular surface patch features," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2015, pp. 4722–4730.

[10] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, 2012, pp. 2879–2886.

[11] K. Pawelczyk and M. Kawulok, "Head pose estimation relying on appearance-based nose region analysis," in *Computer Vision and Graphics*, ser. Lecture Notes in Computer Science, L. Chmielewski, R. Kozera, B.-S. Shin, and K. Wojciechowski, Eds. Springer International Publishing, 2014, vol. 8671, pp. 510–517.

[12] X. Peng, J. Huang, Q. Hu, S. Zhang, and D. Metaxas, "Three-dimensional head pose estimation in-the-wild," in *Automatic Face and Gesture Recognition (FG), IEEE International Conference and Workshops on*, vol. 1, 2015, pp. 1–6.

[13] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel, "Probabilistic temporal head pose estimation using a hierarchical graphical model," in *European Conference on Computer Vision (ECCV).*

[14] K. Chang, W. Bowyer, and P. Flynn, "Multiple nose region matching for 3d face recognition under varying facial expression," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 10, pp. 1695–1700, 2006.

[15] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K. Pal, C. Bhagavatula, and M. Savvides, "Investigating the feasibility of image-based nose biometrics," in *IEEE International Conference on Image Processing (ICIP)*, vol. 2, 2015.

[16] Z. Lubing and W. Han, "Local gradient increasing pattern for facial expression recognition," in *Image Processing (ICIP), IEEE International Conference on*, 2012, pp. 2601–2604.

[17] H. Nam and B. Han, "Learning multi-domain convolutional neural networks for visual tracking," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2016.

[18] A. Abaza, M. A. Harrison, T. Bourlai, and A. Ross, "Design and evaluation of photometric image quality measures for effective face recognition," *IET Biometrics*, vol. 3, no. 4, pp. 314–324, 2014.

[19] 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 *Biometrics: Theory, Applications and Systems (BTAS), IEEE International Conference on*, 2013, pp. 1–8.

[20] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[21] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *Automatic Face and Gesture Recognition (FG) – Net Workshop on Visual Observation of Deictic Gestures*, vol. 6, 2004.

[22] M. Demirkus, J. Clark, and T. Arbel, "Robust semi-automatic head pose labeling for real-world face video sequences," *Multimedia Tools and Applications*, vol. 70, no. 1, pp. 495–523, 2014.

[23] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Computer Vision Workshops (ICCV Workshops), IEEE International Conference on.* IEEE, 2011, pp. 2106–2112.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems 28.* Curran Associates, Inc., 2015, pp. 91–99.

[25] R. Girshick, "Fast r-cnn," in *International Conference on Computer Vision (ICCV)*, 2015.

[26] T. Ojala, M. Pietikainen, and D. Harwood, "Performance evaluation of texture measures with classification based on kullback discrimination of distributions," in *Computer Vision & Image Processing. (IAPR), International Conference on*, vol. 1, 1994, pp. 582–585 vol.1.

[27] B. Jun and D. Kim, "Robust face detection using local gradient patterns and evidence accumulation," *Pattern Recognition*, vol. 45, no. 9, pp. 3304 – 3316, 2012, pattern Recognition and Image Analysis (IbPRIA), Iberian Conference on.

[28] F. H. B. Zavan, "Nose pose estimation in the wild and its applications on nose tracking and 3d face alignment," Master's thesis, UFPR, 2016.

[29] 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 *Conference on Graphics, Patterns and Images (SIBGRAPI) – Workshop on Face Processing Applications*, 2016.

[30] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 7, pp. 673–689, 1996.

[31] F. H. B. Zavan, A. C. P. Nascimento, L. P. e Silva, O. R. P. Bellon, and L. Silva, "3d face alignment in the wild: A landmark-free, nose-based approach," in *European Conference on Computer Vision (ECCV) – Workshop on 3D Face Alignment in the Wild*, 2016, pp. 581–589.

[32] 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 *International Conference on Computer Vision(ICCV), IEEE – Workshops*, 2015, pp. 1003–1011.

[33] 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 *Conference on Graphics, Patterns and Images (SIBGRAPI) – Workshop on Face Processing Applications*, 2016.

[34] L. A. Jeni, S. Tulyakov, L. Yin, N. Sebe, and J. F. Cohn, "The first 3d face alignment in the wild (3dfaw) challenge," in *European Conference on Computer Vision (ECCV) – Workshops*, 2016, pp. 511–520.

[35] L. P. Silva, "Rastreamento facial e refinamento de pontos fiduciais 3d baseado na região do nariz em ambientes não controlados," Master's thesis, UFPR, 2017.

[36] J. C. Batista, V. Albiero, O. R. P. Bellon, and L. Silva, "Aumpnet: simultaneous action units detection and intensity estimation on multipose facial images using a single convolutional neural network," in *Automatic Face and Gesture Recognition (FG), IEEE – FERA Challenge Workshop*, 2017.