Machine Vision and Applications (2021) 32:106
https://doi.org/10.1007/s00138-021-01234-1

SPECIAL ISSUE PAPER

Partitioned iterated function systems by regression models for head pose estimation

Andrea F. Abate[1] · Paola Barra[1] · Chiara Pero[1] · Maurizio Tucci[1]

Received: 8 February 2021 / Revised: 27 May 2021 / Accepted: 15 July 2021 / Published online: 4 August 2021

Abstract
Head pose estimation represents an important computer vision technique in different contexts where image acquisition cannot be controlled by an operator, making face recognition of unknown subjects more accurate and efficient. In this work, starting from partitioned iterated function systems to identify the pose, different regression models are adopted to predict the angular value errors (yaw, pitch and roll axes, respectively). This method combines the fractal image compression characteristics, such as self-similar structures in order to identify similar head rotation, with regression analysis prediction. The experimental evaluation is performed on widely used benchmark datasets, i.e., Biwi and AFLW2000, and the results are compared with many existing state-of-the-art methods, demonstrating the robustness of the proposed fusion approach and excellent performance.

Keywords Head pose estimation · PIFS · Face detection · Regression models

1 Introduction

Head pose estimation (HPE) is a computer vision technique for determining the orientation of a human's head. Head movements represent an important aspect of a subject, providing several characteristics like individual's intentions and attention. In any context where image acquisition cannot be controlled by an operator, automated HPE of an unknown subject makes face recognition much more accurate and efficient [33]. In the last decade, many application systems have been developed based on the estimation of the human head directions and movements, finding applicability in several contexts, such as video surveillance and driving monitoring systems. In the literature, the head rotation movements can be determined in different forms. The usually chosen representation uses the Euler angles. In particular, a 3D vector is obtained, including yaw, pith and roll angles. Figure 1 shows the head pose along the three axes, respectively, x, y and z. Estimating the head movements from 2D images is actually an open and still challenging problem for many applications that require head rotation knowledge. In this paper, we consider a classification method based on fractal self-similarity of images, called HP^{2}IFS [5] and apply four different regression models in order to improve its performance in head pose estimation. We performed an experimental evaluation of this novel method over two well-know datasets: Biwi [11] and AFLW2000 [19]. The article is structured as follows. In Sect. 2, is introduced a literature review of 2D and 3D methods for head pose estimation; Sect. 3 illustrates the HP^{2}IFS method; Sects. 4 and 5 analyze, respectively, the regression methods applied to HP^{2}IFS and the datasets adopted in the experimental phase; Sect. 6 describes the experimental results. Finally, conclusions are showed in Sect. 7.

✉ Chiara Pero
cpero@unisa.it

Andrea F. Abate
abate@unisa.it

Paola Barra
pbarra@unisa.it

Maurizio Tucci
mtucci@unisa.it

[1] Computer Science Department, University of Salerno, Fisciano, Italy

2 Related work

Many HPE algorithms have been proposed over the years. We divide the HPE approaches available in 2D (intensity) or 3D (depth) image work. 3D data imply the use of special sensors and cameras capable of capturing the subject and acquiring its depth, furthermore, for this type of acquisition the operating distance between the camera and the subject

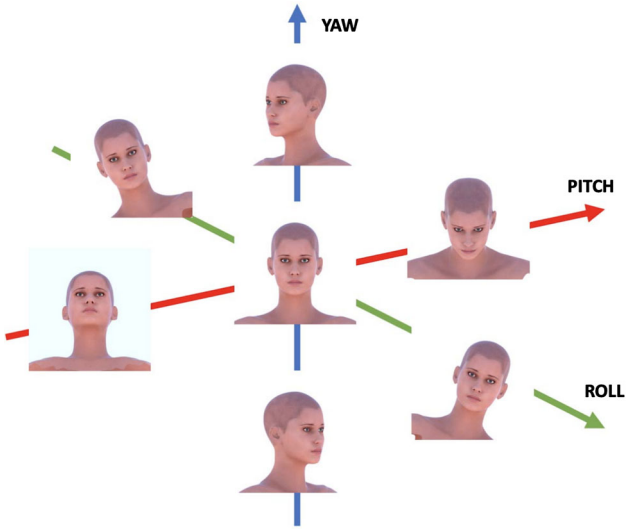

Fig. 1 The head rotations movements represented in yaw, pitch and roll angles

is limited; for these reasons, the use of the above methods in real contexts is very limited and the methods that use 3D images often also use 2D images.

2.1 2D image methods

In the category of approaches working on 2D images, we have many methods involving machine learning techniques in particular with the use of DNN and CNN. The method presented in [28] estimates head pose through a neural network on the Pointing'04 dataset. This dataset contains pitch and yaw information only. FSA-Net [32] is another method that estimates head pose based on the use of a neural network, which is based on regression and aggregation of characteristics. In [31], the authors propose a Coarse-to-Fine strategy using a deep learning approach, jointly training two subnets to classify the frame into four classes and then to estimate the pose via Fine regression. The method in [26] uses the combination of two trained CNNs to identify both the head and the body pose; similarly, the HPE approach in [7] adopts information from video sequence in order to estimate the head orientation through the movement direction analysis of an individual. QuatNet, a multi-regression loss function applied in [17], estimate head rotations with a CNN, using RGB frames without depth information. The work in [21] proposes a whole body estimation method, and it is composed of three steps: (1) in the first step, the person's appearance characteristics are extracted using the HOG technique; (2) the second step updates a classifier with the person's tracking and direction information. Based on the direction in which it walks and the information of the first module, the third step estimates the body orientation, merging the characteristics collected from the previous steps. The authors in [22] analyze the region of the nose, based on its orientation they evaluate the pose of the face. The experiments carried out show that this information has a high discriminatory power to determine the orientation of the head compared to the techniques that are based on the analysis of the entire facial region. In [25,27], through transfer learning two well-known neural networks are used, respectively, Multi-Loss ResNet50 and Hyperface. ResNet50 is used to predict the three face degrees of freedom (yaw, pitch and roll angles, respectively) directly from the image; Hyperface trains a CNN to identify the face region, individuate the facial reference points and estimate the subject pose. In [20], they address the face alignment problem with Kepler that uses Efficient H-CNN Regressors for obtaining iteratively Keypoint Estimation and Pose prediction of unconstrained faces. In [1], the method QuadTree Pitch Yaw and Roll (QT-PYR) is discussed. This approach extracts the 68 landmarks facial points and adopts a QuadTree model to encode the pose through a vector. This vector will be compared to the ground truth to estimate the pose. This method does not make use of neural networks. The papers [2,3] obtain a face pose coding building a Web-Shaped Model through the reference points of the face. In hGLLiM [10], they experiment different classifiers and regression methods, proposing to use a mixture of linear regressions that learns to map high-dimensional feature vectors (extracted from the face bounding boxes) on the head pose angles and the bounding box displacements, so that they are predicted in robust way in the presence of unobservable phenomena. In the method presented in [9], the HPE is formulated as a mixture of linear regression problems. The method maps the HOG-based descriptors extracted from the face bounding boxes to the corresponding head poses. Finally, the authors in [15] address the head pose estimation challenge analyzing low-resolution frames with a large angles range and using chrominance-based functions. These images constitute the input for a linear auto-associative memory, which is calculated for each head pose using a Widrow–Hoff learning rule.

2.2 3D image methods

The majority of the existing solutions operate in 2D images, but 3D imaging has also been exploited here. For example, [12] explores the orientation of a human's head using depth information. The authors, with a statistical model of the face, train a large amount of synthesize and annotated data. The experimental evaluation demonstrates that the method is capable of handling real-world data with non-cooperative subjects, partial occlusions of facial regions and facial expression changes, even if it is only trained on synthetic facial data. In [34], 3DDFA (3D Dense Face Alignment) is proposed, which adapts a dense Morphable 3D model (3DMM) of a face to an image via cascading CNN. In [6], FAN is presented, in which a very large 2D dataset is

synthetically expanded by converting the annotations of the 2D landmarks into 3D and unifying all the existing datasets, leading to the creation of LS3D-W. The method presented in [8] introduces a robust method in the case of variable lighting and rotation. Head pose is estimated from 2D key points drawn in two consecutive frames in the head region and their 3D projection on a simple geometric model. In the automotive field, [24] presents a solution for monitoring the driver's head. By combining 2D and 3D information, head position is estimated and regions of interest identified. This is to detect special driver-related events such as drowsiness or inattention.

3 HP^{2}IFS: partitioned iterated function systems for head pose estimation

The method adopted to estimate an individual's head pose is proposed in [5]. This approach is closely related to fractal image compression and, consequently, to the concept of partitioned iterated function systems (PIFS) [14]. In particular, fractal compression bases its origin on *self-similar* structures, which possess almost the same features at any level of detail they are enlarged. Thus, it is possible to describe and generate fractals using extremely simple recursive deterministic algorithms, gradually producing copies of oneself or portions of oneself at various scaling factors. Fractal compression essentially consists in searching, for the whole image or part of it, the fractal object that is best suited to approximating its information content and in encoding the description of the object associated with the image. Originally used as a lossy image compression algorithm, the HP^{2}IFS approach [5] allows to analyze the self-similarity of two images representing a similar head rotation.

The main steps characterizing the fractal encoding algorithm are the following:

1. Partition the input image into R_i non-overlapping blocks of size $N \times N$ (namely *Range Blocks*).
2. Partition the input image into overlapping D_j blocks of size $2N \times 2N$ (*Domain Blocks*).
3. Determine the self-similar parts within the image, memorizing every possible area in terms of the image itself through *contractive transformations*, i.e., applying various combination of geometrical transformations and luminance factors.

Therefore, iterating a series of affine transformations f_i, the fractal compression algorithm goal is the finding of best matching block for each R block, satisfying the minimum distortion error. These transformations represent the fractal encoding result. So, in HP^{2}IFS method, given an image as input, we identify the face using Viola Jones's algorithm [30]. Then, using a pre-trained regression method [18], the 68 facial landmark points are identified in order to create a facial mask. The resulting mask is encoded using the fractal compression algorithm (see Fig. 2). As above-mentioned, the matrix created by fractal encoding is converted into a pose feature vector which will be compared with the built reference model.

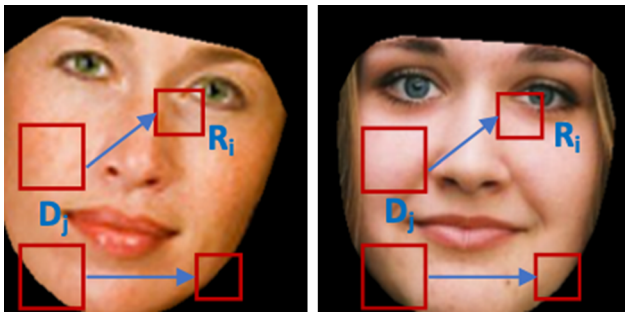

Fig. 2 Fractal encoding process: domain and range blocks

4 HP^{2}IFS: regression models

To estimate the pose of an individual, we use the classification method shown in Fig. 3B) [5], and subsequently, we compare with the regression approach. In particular, the resulting array from fractal encoding is compared with a reference model using the Hamming distance [16]. The reference model is obtained from a part of the dataset involved in the tests.

Regression analysis is a predictive modeling technique in which the target variable to be estimated is continuous. By definition, regression represents the learning process of a target function f that maps each attribute x to continuous output [13]. So, the goal is to find the target function that is able to adapt to the input with the minimum error. In this work, starting from HP^{2}IFS approach to identify the pose, we adopt 4 different regression models to perform the results, for yaw, pitch and roll angles. This procedure is illustrated in Fig. 3C). Further details are present in the following subsections.

4.1 Linear regression

Linear regression (LR) represents the simplest form of regression [4]. The relationship between dependent and independent variables is assumed to be linear. In Eq. 1, y represents the dependent variable to be estimated, x and ϵ are, respectively, the independent variable and the error term. β is the regression coefficient.

$$y = \beta x + \epsilon \tag{1}$$

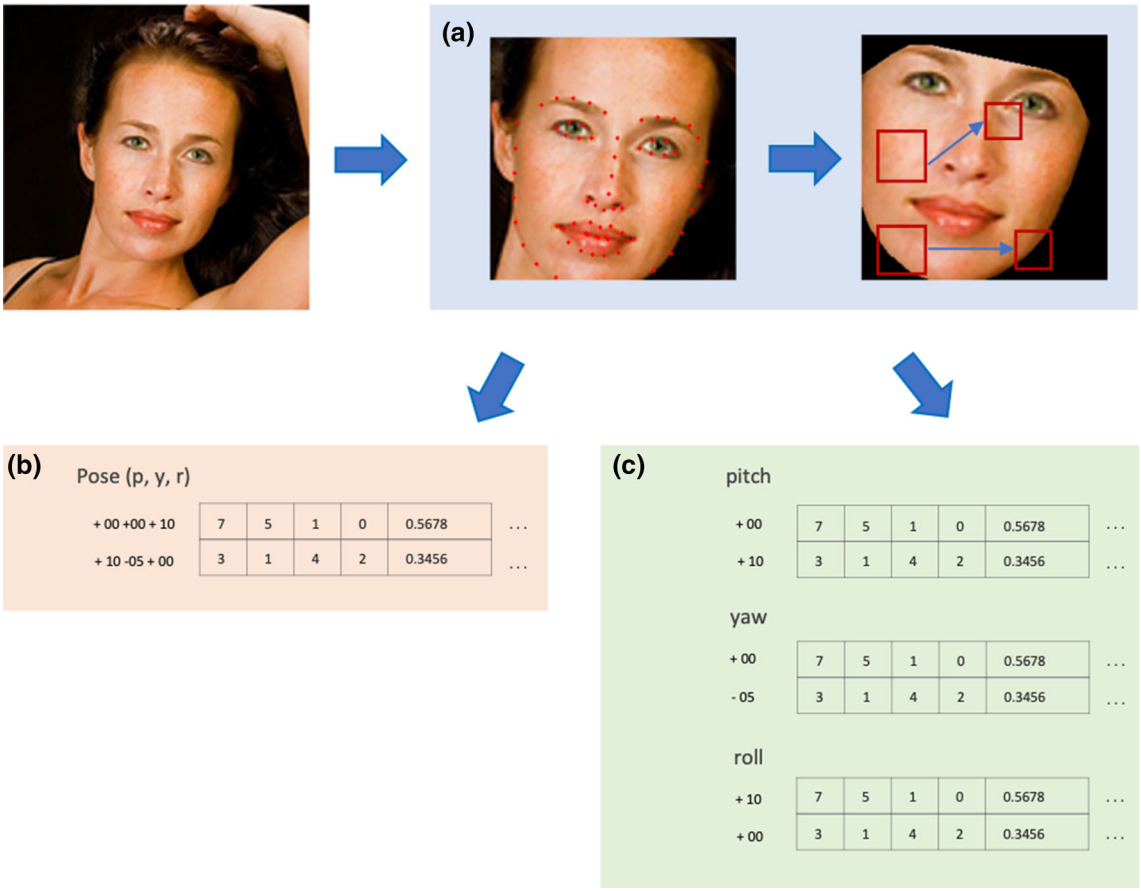

Fig. 3 Framework of the proposed method: **A** HP^{2}IFS approach; **B** classification; **C** regression

A relationship between variables of interest does not necessarily imply that one variable is the cause of the other, but that there is a significant association between the two variables.

4.2 Bayesian ridge regression

Ridge regression, also known as Tikhonov regularization, is a classical regularization technique of Linear regression [29]. This model estimate has a Bayesian interpretation. In particular, adopting a fully probabilistic model, in which the prior of the coefficients are given by a spherical Gaussian, it is possible to obtain a Ridge regression using a Bayesian view (see Eq. 2).

$$p(w|\lambda) = \mathcal{N}(w|0, \lambda^{-1}, \mathbf{I}_p) \tag{2}$$

4.3 Logistic regression

Logistic regression (LgR), also called as Logit model, is a nonlinear regression model used when the dependent variable is dichotomous. LgR through statistical methods allows to generate a result which represents a probability that a given input value belongs to a specific class. The goal is to establish the probability with which an observation can generate one or the other value of the dependent variable [23]. Eq. 3 refers to Logit model:

$$y = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \tag{3}$$

4.4 Lasso regression

Lasso regression, acronym of Least Absolute Shrinkage and Selection operator, is a regularized version of Linear regression [13]. It adopts the $L1$ penalty in the objective function. So, the optimization objective is expressed by Eq. 4:

$$\min \frac{1}{2n_{\text{samples}}} ||y - Xw||_2^2 + \alpha||w||_1 \tag{4}$$

Lasso regression performs a selection of the independent variables, bringing the remaining ones to zero through an appropriate value of the associated weight, and generating a sparse model.

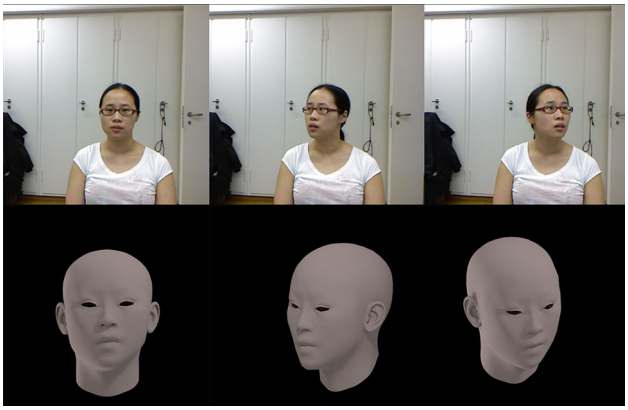

Fig. 4 Some RGB and depth frames from the Biwi dataset with different head-poses

Fig. 5 Samples from the AFLW2000 dataset with different head-poses

5 Datasets

The following datasets were used for experimentation and comparison with the state-of-the-art: *Biwi* dataset [11] and *AFLW2000* dataset [19].

5.1 Biwi dataset

Biwi Kinect Head Pose Database [11] contains RGB-D images of 20 different people (6 females and 14 males) with a total of over 15,000 frames. For each subject, it includes a file with extension .obj with the three-dimensional model of the head of the subject. In Fig. 4, there are five video frames of subject 01 (top) and corresponding depth frames of the same subject (bottom). For 10 subjects, 3D models of the individuals' heads were processed with the Blender graphics engine to obtain 2223 different poses from each 3D subject. For each subject, there are all the possible combinations of pose in terms of pitch, yaw and roll angles (13 variations in pitch, 19 in yaw and 9 in roll) with steps of 5° each. Through this procedure, it was possible to annotate each frame with pitch, yaw and roll, in order to use the figures as a ground truth for experiments.

5.2 AFLW2000 dataset

The AFLW2000 dataset [19] provides the first 2000 images of the Annotated Face Landmarks in the Wild (AFLW) dataset, extract from Flickr social network. In AFLW, the faces depicted are annotated with the pose of the face in the degrees of yaw, pitch and roll. These faces have random poses and different ages, ages, facial expressions, environmental conditions, etc. In Fig. 5, there are some image extracted from the AFLW2000 dataset.

Table 1 Results on the subsets of Biwi applying HP^{2}IFS-BRR model

Subset	Yaw	Pitch	Roll	Overall MAE
1	7.06	7.22	3.47	5.91
2	5.72	4.6	2.39	4.23
3	7.27	6.72	11.66	8.55
4	6.38	5.15	2.65	4.72
5	5.96	4.56	3.06	4.52
6	6.18	5.06	2.83	4.69
7	6.82	6.1	3.06	5.32
8	7.13	4.9	2.7	4.91
9	6.23	4.78	3.37	4.79
10	7.2	5.58	2.83	5.20
Mean	6.59	5.46	3.80	5.28

6 Experimental results

As introduced in Sect. 5, the experiments were performed on BIWI and AFLW2000 datasets. Biwi dataset provides 10 identity with several images each. Consequently, we created our model applying a one-left-out technique. In particular, we performed 10 different experiments set using in turn 1 subject as a tester and the others as a model. Each individual has a wide range of poses that cover the angular variation over the three degrees of freedom. Table 1 shows the results obtained for each subject tested in terms of MAE, applying the regression models mentioned in Sect. 4. The presence of large variations in errors between an experimental subset and the other demonstrates the geometrical difference between the faces of the various subjects.

For AFLW2000 database, the 70% of the frames randomly selected were used to create the model reference and the remaining 30% were adopted for the test. The results obtained by the combination of HP^{2}IFS method and regression models were analyzed through the *Mean Absolute Error* (MAE), a

Table 2 MAE (degrees) of yaw, pitch and roll on Biwi database

Method	Yaw	Pitch	Roll	MAE
Coarse-to-Fine [31]	4.76	5.48	4.29	4.84
FSA-Net [32]	4.27	4.96	2.76	3.99
hGLLiM [10]	6.06	7.65	5.62	6.44
Multi-Loss ResNet50 [27]	5.17	6.97	3.39	5.17
QT-PYR [1]	5.41	12.80	6.33	8.18
QuatNet [17]	4.01	5.49	2.93	4.14
HP^{2}IFS [5]	**4.05**	6.23	**3.30**	**4.52**
HP^{2}IFS-LR	6.57	5.47	3.80	5.28
HP^{2}IFS-BRR	6.59	**5.46**	3.80	5.28
HP^{2}IFS-LgR	9.73	5.82	6.22	7.86
HP^{2}IFS-LsR	6.58	5.29	3.80	5.28

The bold values are represent the comparison between HP2IFS method and the novel approach described in this paper

Table 3 MAE (degrees) of yaw, pitch and roll on AFLW2000 database

Method	Yaw	Pitch	Roll	MAE
3DDFA [34]	5.40	8.53	8.25	7.39
FAN [6]	6.35	12.27	8.71	9.11
Hyperface [25]	7.61	6.13	3.92	5.89
Kepler [20]	6.45	5.85	8.75	7.01
Multi-Loss ResNet50 [27]	6.47	6.55	5.43	6.15
QT-PYR [1]	7.6	7.6	7.17	7.45
QuatNet [17]	3.97	5.61	3.92	4.50
HP^{2}IFS [5]	**6.28**	7.46	5.53	6.42
HP^{2}IFS-LR	6.71	6.90	4.48	6.03
HP^{2}IFS-BRR	6.59	7	5.19	6.26
HP^{2}IFS-LgR	8.16	7.71	5.86	7.24
HP^{2}IFS-LsR	6.70	**6.90**	**4.48**	**6.02**

The bold values are represent the comparison between HP2IFS method and the novel approach described in this paper

performance index commonly used in HPE evaluation. MAE measures the average over the absolute differences between the predicted values (in this case, the predicted poses) and the actual observation (i.e the ground truth poses), as indicated in Eq. 5:

$$\text{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y_j}| \tag{5}$$

where y_j is the angular value of true pose and $\hat{y_j}$ is the angular value of predicted pose. The comparison with existing literature review, described in Sect. 2, is reported in the following tables. Our HP^{2}IFS regression methods on Biwi and AFLW2000 datasets are showed, respectively, in Tables 2 and 3. The proposed fusion approach includes four types of regression: Linear (HP^{2}IFS-LR), Bayesian Ridge (HP^{2}IFS-BRR), Logistic (HP^{2}IFS-LgR) and Lasso (HP^{2}IFS-LsR).

All values reporting in the tables represent the MAE for each of the three angular poses, including an overall MAE along the three axes.

Table 2 shows the results performed on Biwi dataset and compared with other state-of-art approaches. In Bayesian Ridge regression model, it is possible to observe the roll angular error and the overall MAE similar to HP^{2}IFS classification method, and the pitch angular error better than some other deep learning-based approaches. HP^{2}IFS yaw angular error represents the only exception.

Table 3 reports the comparison results on AFLW2000 database. The Lasso regression model provides lowest MAE value respect to all other state-of-the-art methods, including pitch and roll angular errors. Very few exceptions, as for the Biwi dataset in Table 2, are related to methods that use the neural networks. It can also be noted that HP^{2}IFS-LsR yaw angular error value is very close to HP^{2}IFS yaw error. Finally, Figs. 6 and 7 illustrate, respectively, the error distribution in terms of percentage of tested images using Bayesian Ridge regression model (BRR) and Lasso regression model (LsR) on Biwi and AFLW2000 datasets, thus showing a similar trend anticipated by the results (see Tables 2, 3). In particular, for BIWI dataset 90% of the poses has error less than 20° and maximum error equal to 60° for yaw (see Fig. 6). For AFLW2000 benchmark, 90% of the images have an error less than 15° and maximum error equal to 35° for yaw, as can be seen in Fig. 7. Since x shows the pose estimation error in degrees, we can see that for yaw there is a higher percentage of poses and a higher error; this is because the images have pose variations with a wider range in yaw.

The total time to perform the pre-processing phase on an image with size 256×256 is 0.06 s, including the face detection and localization, the landmark prediction and, finally, the mask creation process. All the experiments are performed on a MacBook Pro 2.6 GHz Intel Core i7 6 core 16 GB 2667 MHz DDR4 Intel UHD Graphics 630 1536 MB, with Python 3.6.8.

7 Conclusions

In this work, four different regression methods combined with HP^{2}IFS approach are analyzed to estimate an individual's head pose. In particular, HP^{2}IFS regression method merges fractal image compression self-similarity properties with regression models prediction, thus identifying similar head rotations. The experiments carried out on widely-used benchmark datasets including Biwi and AFLW2000 are compared with many state-of-the-art approaches, demonstrating excellent performance and obtaining accurate angular values

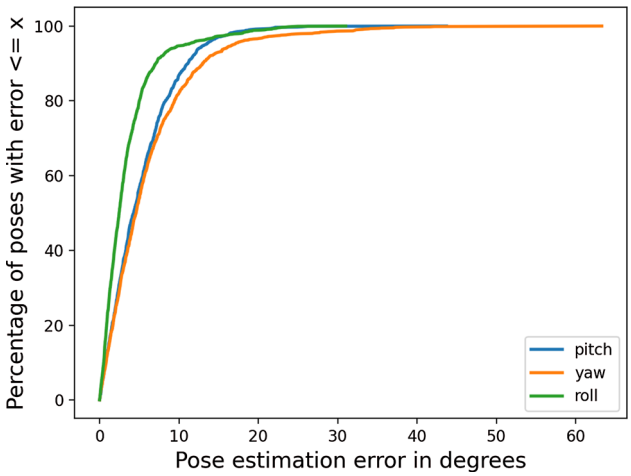

Fig. 6 Errors on Biwi dataset respect to the tested images (%) in HP^{2}IFS-BRR

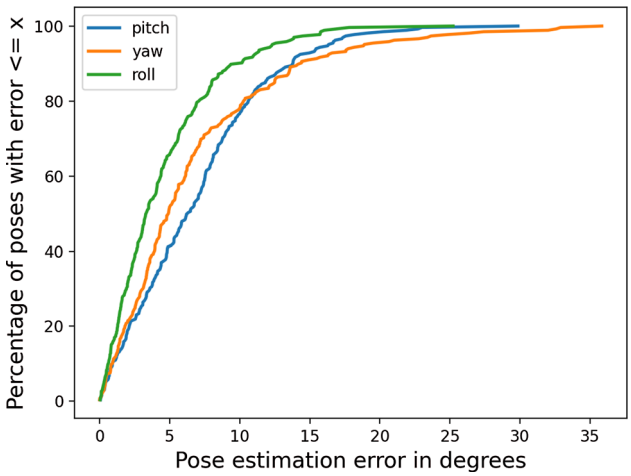

Fig. 7 Errors on AFLW2000 dataset respect to the tested images (%) in HP^{2}IFS-LsR

along the three axes, i.e., for yaw, pitch and roll. The proposed fusion methodology is superior to other deep-learning-based methods, and it also requires no training phase.

Funding Open access funding provided by Universitá degli Studi di Salerno within the CRUI-CARE Agreement.

References

1. Abate, A.F., Barra, P., Bisogni, C., Nappi, M., Ricciardi, S.: Near real-time three axis head pose estimation without training. IEEE Access **7**, 64256–64265 (2019)
2. Abate, A.F., Barra, P., Pero, C., Tucci, M.: Head pose estimation by regression algorithm. Pattern Recognit. Lett. **140**, 179–185 (2020)
3. Barra, P., Barra, S., Bisogni, C., De Marsico, M., Nappi, M.: Web-shaped model for head pose estimation: an approach for best exemplar selection. IEEE Trans. Image Process. **29**, 5457–5468 (2020). https://doi.org/10.1109/TIP.2020.2984373
4. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Berlin (2006)
5. Bisogni, C., Nappi, M., Pero, C., Ricciardi, S.: Hp2ifs: head pose estimation exploiting partitioned iterated function systems. In: 25th International Conference on Pattern Recognition (ICPR2020)
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). In: International Conference on Computer Vision (2017)
7. Chamveha, I., Sugano, Y., Sugimura, D., Siriteerakul, T., Okabe, T., Sato, Y., Sugimoto, A.: Appearance-based head pose estimation with scene-specific adaptation. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1713–1720 (2011). https://doi.org/10.1109/ICCVW.2011.6130456
8. Díaz Barros, J.M., Mirbach, B., Garcia, F., Varanasi, K., Stricker, D.: Real-time head pose estimation by tracking and detection of keypoints and facial landmarks. In: Bechmann, D., Chessa, M., Cláudio, A.P., Imai, F., Kerren, A., Richard, P., Telea, A., Tremeau, A. (eds.) Computer Vision, Imaging and Computer Graphics Theory and Applications, pp. 326–349. Springer, Cham (2019)
9. Drouard, V., Ba, S., Evangelidis, G., Deleforge, A., Horaud, R.: Head pose estimation via probabilistic high-dimensional regression. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 4624–4628. IEEE (2015)
10. Drouard, V., Horaud, R., Deleforge, A., Ba, S., Evangelidis, G.: Robust head-pose estimation based on partially-latent mixture of linear regressions. IEEE Trans. Image Process. **26**(3), 1428–1440 (2017)
11. Fanelli, G., Dantone, M., Gall, J., Fossati, A., Van Gool, L.: Random forests for real time 3D face analysis. Int. J. Comput. Vis. **101**(3), 437–458 (2013)
12. Fanelli, G., Gall, J., Van Gool, L.: Real time head pose estimation with random regression forests. In: CVPR 2011, pp. 617–624 (2011). https://doi.org/10.1109/CVPR.2011.5995458
13. Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., James, S.: Machine learning for cultural heritage: a survey. Pattern Recognit. Lett. **133**, 102–108 (2020)
14. Fisher, Y.: Fractal image compression. In: Fractals in Engineering—Proceedings of the Conference on Fractals in Engineering, vol. 94, p. 165. World Scientific (1995)
15. Gourier, N., Maisonnasse, J., Hall, D., Crowley, J.L.: Head pose estimation on low resolution images. In: International Evaluation Workshop on Classification of Events, Activities and Relationships, pp. 270–280. Springer (2006)
16. Hamming, R.W.: Error detecting and error correcting codes. Bell Syst. Tech. J. **29**(2), 147–160 (1950)
17. Hsu, H.W., Wu, T.Y., Wan, S., Wong, W.H., Lee, C.Y.: Quatnet: quaternion-based head pose estimation with multiregression loss. IEEE Trans. Multimed. **21**(4), 1035–1046 (2018)
18. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)

19. 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. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 2144–2151. IEEE (2011)
20. Kumar, A., Alavi, A., Chellappa, R.: Kepler: keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 258–265. IEEE (2017)
21. Liu, H., Ma, L.: Online person orientation estimation based on classifier update. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 1568–1572 (2015). https://doi.org/10.1109/ICIP.2015.7351064
22. Pawelczyk, K., Kawulok, M.: Head pose estimation relying on appearance-based nose region analysis. In: Chmielewski, L.J., Kozera, R., Shin, B.S., Wojciechowski, K. (eds.) Computer Vision and Graphics, pp. 510–517. Springer International Publishing, Cham (2014)
23. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
24. Peláez, C., García, F., de la Escalera, A., Armingol, J.M.: Driver monitoring based on low-cost 3-D sensors. IEEE Trans. Intell. Transp. Syst. **15**(4), 1855–1860 (2014). https://doi.org/10.1109/TITS.2014.2332613
25. Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern Anal. Mach. Intell. **41**(1), 121–135 (2019). https://doi.org/10.1109/TPAMI.2017.2781233
26. Raza, M., Chen, Z., Rehman, S.U., Wang, P., Bao, P.: Appearance based pedestrians' head pose and body orientation estimation using deep learning. Neurocomputing **272**, 647–659 (2018)
27. Ruiz, N., Chong, E., Rehg, J.M.: Fine-grained head pose estimation without keypoints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 2074–2083 (2018)
28. Stiefelhagen, R.: Estimating head pose with neural networks-results on the pointing04 ICPR workshop evaluation data. In: Proceedings of Pointing 2004 Workshop: Visual Observation of Deictic Gestures, vol. 1, pp. 21–24 (2004)
29. Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. **1**(Jun), 211–244 (2001)
30. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2001, vol. 1, p. I. IEEE (2001)
31. Wang, Y., Liang, W., Shen, J., Jia, Y., Yu, L.F.: A deep coarse-to-fine network for head pose estimation from synthetic data. Pattern Recognit. **94**, 196–206 (2019)
32. Yang, T.Y., Chen, Y.T., Lin, Y.Y., Chuang, Y.Y.: FSA-net: learning fine-grained structure aggregation for head pose estimation from a single image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1087–1096 (2019)
33. Zhang, F., Zhang, T., Mao, Q., Xu, C.: Joint pose and expression modeling for facial expression recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3359–3368 (2018)
34. Zhu, X., Liu, X., Lei, Z., Li, S.Z.: Face alignment in full pose range: a 3D total solution. IEEE Trans. Pattern Anal. Mach. Intell. **41**(1), 78–92 (2019). https://doi.org/10.1109/TPAMI.2017.2778152

Andrea F. Abate currently serves as an Associate Professor with the University of Salerno from 2006, where he is team leader of the Virtual Reality Laboratory. Dr. Abate is a member of the IEEE Haptics Technical Committee and a member of the the International Association for Pattern Recognition. His current research interests include multibiometric systems, virtual/augmented/mixed reality, haptics and human–computer interaction. He has authored many scientific papers published in scientific journals and proceedings of refereed international conferences and co-edited one book. He currently serves as Associate Editor for Pattern Recognition Letters and IEEE Access.

Paola Barra received PhD degree in Computer Science from University of Salerno, the M.S. degree in Business Informatics from University of Pisa and the B.S. degree in Computer Science from University of Salerno. She is currently research fellow at University of Rome La Sapienza. Her research interests include Machine Learning technics in facial and gait recognition, image processing and video games development. She is member of IEEE and GIRPR/IAPR.

Chiara Pero received the B.S. and M.S. (cum laude) degrees in Computer Science from the University of Salerno, Italy, in 2016 and 2018, respectively. She is currently pursuing the Ph.D.degree in Computer Science with the Biometric and Image Processing Laboratory (BIPLAB), University of Salerno. Her research interests include Machine Learning technics in facial recognition, image processing, cancelable biometrics and behavioral profiling.

Maurizio Tucci is currently a Full Professor in Computer Science at the University of Salerno, Dipartimento di Informatica. He served as the responsible for the Graduation Programs in Computer Science from 2007 to 2010. Formerly, he was the Director of the Department of Mathematics and Computer Science of the University of Salerno, from 2000 to 2006. He has about 30 years of research activity covering various aspects of formal models for the development of visual languages and systems, human-computer interaction, image indexing and retrieval, software engineering, geographical information systems. His current research focuses on statistical data visualization, usability evaluation of GUIs, and mobile interactive systems and service development.